2019-12-17

Modulating MCM Levels Causes Differential Loading at Origins of Replication and Changes Replication Timing

Livio Dukaj
University of Massachusetts Medical School

MODULATING MCM LEVELS CAUSES DIFFERENTIAL LOADING AT ORIGINS OF REPLICATION AND CHANGES REPLICATION TIMING

A Dissertation Presented

By

LIVIO DUKAJ

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences, Worcester
In partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 17, 2019

Biochemistry and Molecular Pharmacology

SIGNATURE PAGE

MODULATING LEVELS OF MCM CAUSES DIFFERENTIAL LOADING AT ORIGINS OF REPLICATION AND CHANGES REPLICATION TIMING

A Dissertation Presented By

LIVIO DUKAJ

This work was undertaken in the Graduate School of Biomedical Sciences
(Biochemistry and Molecular Pharmacology)

Under the mentorship of:

Nicholas Rhind, PhD, Thesis Advisor

Brian Kelch, PhD, Member of Committee

David Grunwald, PhD, Member of Committee

William Kobertz, PhD, Member of Committee

David MacAlpine, PhD, External Member of Committee

Sharon Cantor, PhD, Chair of Committee

Mary Ellen Lane, PhD
Dean of the Graduate School of Biomedical Sciences

December 17, 2019

ACKNOWLEDGEMENTS

I would like to start off by thanking Nick for his support and guidance throughout these years. His contagious enthusiasm and curiosity for science has made it possible to navigate graduate school through all its peaks and valleys. Thank you to Rhind lab members, current and from the past, for their friendship and collegiality throughout the years. Also thanks to Karen (and previous administrative assistants) whose logistical help has made doing the work as easy as possible.

Thank you to my TRAC committee - Sharon, David, and Brian – for tons of helpful guidance throughout my many projects and for always encouraging me to think about the big picture. Thank you to Dave M. for making the trip from far away and taking part in my DEC, as well as Bill for joining to make the defense a possibility.

Thanks to the Rando lab for their generosity and expertise. Special thanks to Nils for his help and advice in many of the technical challenges in the wet and dry lab parts of my thesis work.

Thanks to the Munson and Kelch labs, and all BMP labs in general, for always being helpful with gear and technical advice. Thanks also to Moore lab for sharing their TIRF microscope and other instruments.

Special thanks to the Walter Lab for their hospitality and generosity during the early years of my graduate career. Thanks also to the Bell lab for providing many of the strains and antibodies used in this thesis work. To Dima from the Nieduszysnki for help with replication timing scripts. To the Benanti lab for sharing their instruments and as well as their advice during the early years of shared lab meetings.

Lastly, the biggest thank you to my family, friends, and loved ones for their unwavering support throughout this journey. I couldn't have done it without you.

ABSTRACT

DNA replication is a highly complex part of cell metabolism that ensures safe propagation of the genome through tight regulation of the expression, localization, and activity of a large number of factors. Replication starts from distinct sites in the genome and initiation events are temporally ordered in a manner that is, on average, highly reproducible across cell populations. The specific order with which different parts of the genome are replicated has been proposed to be important to processes such as gene expression, cell differentiation, development, and genome evolution. Nevertheless, the fundamental mechanisms that are responsible for establishing these timing programs remain elusive.

Unlike in higher eukaryotes, DNA replication in budding yeast initiates at sequence-specific loci called origins of replication. The timing of initiation at these loci is determined by the activation of the main replicative helicase Minichromosome Maintenance (MCM) complex. Recent results have placed MCM in a key role in establishing a replication timing program that is reproducible but arises from stochastic activation of origins, as has been observed in yeast and higher eukaryotes. One particular model posits that the loading of multiple MCMs at individual origins increases the chances that origins will be activated earlier in S phase by a limited amount of initiation factors.

To further test this model, we set out to examine the consequences of modulating MCM levels in budding yeast in order to ascertain their effects on the dynamics of helicase loading during G1 and subsequent replication timing. Overexpression of MCM2-7 had no effects on cell viability, cell cycle progression, MCM abundance at origins, or replication timing. On the other hand, depletion of Mcm4, one of the six obligate components of the MCM helicase, caused reduced viability, slower progression through S phase, and increased sensitivity to replication stress. Importantly, Mcm4 depletion led to differential reduction in MCM loading at origins during G1, with low MCM origins being disproportionately affected by reduced MCM pools. Finally, reduced MCM loading at origins of replication led to delayed replication during S phase. Our data support a model where the loading activity of origins, controlled by their ability to recruit ORC and compete for MCMs, determines the number of helicases loaded, which in turn has strong implications for replication timing.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

MCM	Minichromosome Maintenance Complex
DH	Double Hexamer
pre-RC	Pre Replication Complex
ORC	Origin Recognition Complex
ARS	Autonomously Replicating Sequence
ACS	ARS Consensus Site
ssDNA	Single Stranded DNA
dsDNA	Double Stranded DNA
HU	Hydroxyurea
NFR	Nucleosome Free Region
rDNA	Ribosomal DNA
GFP	Green Fluorescent Protein
CDK	Cyclin Dependent Kinase
NLS	Nuclear Import Signal
NES	Nuclear Export Signal

CHAPTER I: INTRODUCTION

1. Project Overview

In order to ensure successful duplication of the genome, eukaryotic cells tightly regulate the localization, expression, and activity of many proteins involved in DNA replication. Of particular importance, the establishment of the complexes necessary for replication is tightly and redundantly separated from their activation. This ensures that DNA is replicated once and only once per cell cycle, which is crucial in order to avoid chromosomal aberrations and cell death.

One of the key factors involved in DNA replication is the minichromosome maintenance (MCM) complex. This hexamer is an essential protein that has been shown to function as the main replicative helicase that unzips DNA during replication. Its association with DNA is tightly regulated in all eukaryotes and changes in its expression levels have been implicated in numerous diseased states. In addition to its role in active replication, by virtue of its essential role and the initiation mechanism in eukaryotes, MCM localization also determines where replication initiates. This introduction lays the background for a thesis project that explores the physiological and mechanistic effects of modulating MCM levels. In addition, the experiments in this project further test models for how MCM is loaded onto DNA and how it may be involved in the establishment of the DNA replication program in budding yeast.

2. *Saccharomyces cerevisiae* as a Model Organism to Study DNA Replication

Humans have used yeast's ability to ferment sugar for millennia. This process has been used in baking, brewing, and winemaking by numerous cultures spanning the Earth. Though its first uses in the laboratory date back over a century, it wasn't until the 1930s that genetic work on yeast started (Mortimer 2000). Yeast possesses many qualities that make it a great model organism, including it's ease of use and cost of growth, the ability to visually determine the phase of the cell cycle based on morphology, and the ease with which it's genome can be modified through homologous recombination (Williamson 1965; Hartwell et al. 1974; Jasin & Rothstein 2013). A very important feat was achieved in 1996 when *Saccharomyces cerevisiae* became the first eukaryotic organism to have its genome sequenced (Goffeau et al. 1996). Since then, analysis of the yeast genome and its comparison with the human genome has shown that over 2500 genes, or one-third of the yeast genome, have a human ortholog (O'Brien et al. 2005). In addition, nearly a quarter of disease-causing genes in humans have a yeast ortholog (Forslund et al. 2011). Altogether, these characteristics make S. cerevisiae an excellent model organism to study various cellular processes.

The pathway of DNA replication is exceptionally well-conserved between yeast and humans. Virtually every major component involved in the licensing,

initiation, elongation, and termination pathways in yeast has a well-conserved homolog in humans. Particularly, the heterohexameric MCM helicase and the components required for loading it onto DNA before the start of replication are very well conserved in sequence and function (Forsburg 2004; Moiseeva & Bakkenist 2018).

3. Overview of DNA Replication in *S. cerevisiae*

DNA replication is a tightly regulated aspect of cell metabolism. From start to finish, each step has implications for how the next will function. For example, where and how origins are licensed has implications for where and when replication initiates. Where replication initiates has impactions for where it terminates. The number of replication initiations occurring has implications for how far each of them will elongate. Therefore, to thoroughly understand how any one of the steps works, it is useful to know how each of the steps involved in DNA replication are affected by one another.

3.1 Licensing

The first step in DNA replication is the licensing of origins, which results in two high-salt-resistant MCM helicases stably loaded onto double-stranded DNA to form a pre-replication complex (pre-RC) (Yardimci & Walter 2014).

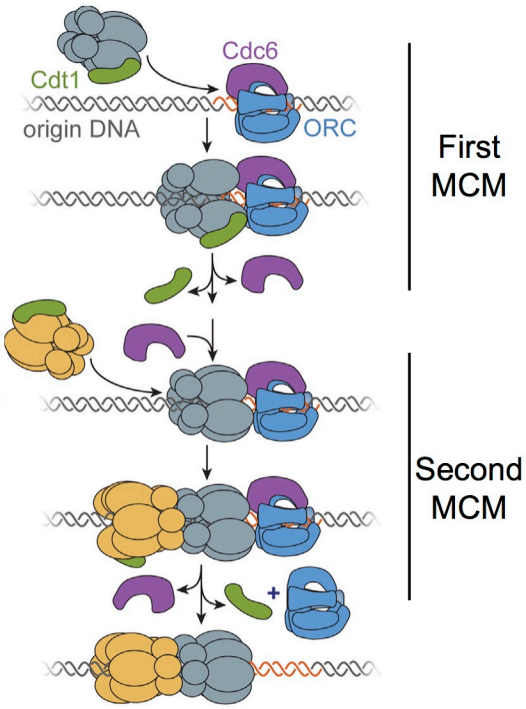


Figure 1.1: Licensing of origins of replication in budding yeast.
The first Cdt1-bound MCM2-7 helicase is recruited by ORC-Cdc6 and loaded onto the origin. Following release of the Cdt1 and Cdc6, another Cdc6 molecule binds ORC and a second Cdt1-MCM2-7 is recruited and loaded onto the origin, forming a bidirectionally-oriented MCM2-7 double hexamer. (Adapted from Ticau et al. Cell, 2015)

From recent single-molecule experiments, the fine molecular details of licensing have largely been elucidated. First, the heterohexameric origin recognition complex (ORC) binds the conserved 17 base-pair ARS consensus sequence (ACS). Cdc6 then binds ORC and together the complex recruits a Cdt1-bound and open MCM2-7 ring. Following loading and closure of the MCM2-7 ring, Cdt1 is released, followed by Cdc6 (Figure 1.1) (Ticau et al. 2015, 2017).

The biochemistry described above depicts the loading of one MCM hexamer. However, origins are licensed when two MCM hexamers (aka a double hexamer)

are loaded in a head-to-head orientation, with the N termini touching and protecting a total of ~68bp of DNA (Evrin et al. 2009; Remus et al. 2009). One double hexamer is the minimal number required for origin licensing, but multiple double hexamers may also be loaded.

A point of discussion remains in the field in how loading of the second MCM hexamer occurs to form a full double hexamer ready for activation. Data from single-molecule experiments suggests that same ORC complex that loads the first MCM hexamer is bound by another Cdc6 molecule and recruits a second Cdt1-MCM2-7 open ring. Although the loading of the first MCM is relies on ORC, this work indicates that the loading of the second MCM helicase depends largely on interactions between the first, loaded MCM and a second Cdt1-bound open hexamer (Ticau et al. 2015). Overall, ATP binding by ORC, Cdc6, and MCM is required for the formation of this stably loaded double-hexamer (Bell & Stillman 1992; Coster et al. 2014; Kang et al. 2014).

Another model for double hexamer loading, using an assay measuring bulk loading of MCM, found that two ORC complexes are needed for efficient loading of double hexamers (Coster & Diffley 2017). The efficiency of MCM loading in these experiments was dependent on the presence of two ORC binding sites, generally a high affinity site and a low affinity site. There seemed to be no required distance between these binding sites for efficient MCM loading up to the 400 bp tested, with the highest amount of loading occurring when the ORC

binding sites were 70 base pairs apart. However, the orientation of the ORC binding sites had strong implications for MCM loading, with sites oriented in a head-to-head orientation being necessary for efficient loading. In addition, this work suggested that stably-loaded MCM hexamers resulting from each ORC binding site likely slide along DNA to form the head-to-head double hexamer. Also from this work, assays monitoring yeast survival and plasmid loss using the same constructs indicate that the presence and orientation of these ORC binding sites is crucial for origin function in vivo. Lastly, more bulk analysis shows that a second ORC binding site, such as the B2 element of ARS1, may be necessary for efficient double hexamer loading in templates that are chromatinized or are artificially modified with flanking roadblocks, supporting a mechanism where sliding helicase-loading intermediates are important for origin function (Warner et al. 2017).

Both of the in vitro assays mentioned above used a similar purification procedure for ORC, Cdc6, MCM, and Cdt1. One notable difference is the concentration of proteins used in the assays, with the single molecule experiments using a concentration of proteins an order of magnitude lower. Nevertheless, the single molecule experiments seem to point to the ability of a single ORC to load a double hexamer of MCM. Whether this is the prevalent mechanism in vivo remains to be seen. However, given the in vivo data about origin function, the involvement of a second ORC for proper origin function seems likely.

Most recently, structural studies using cryo electron microscopy have revealed snapshots of the licensing program that may reconcile the single ORC and double ORC models. From these studies, the second MCM hexamer is loaded via a quasi-symmetrical mechanism involving ORC binding to the B2 element in an inverted orientation relative to the ACS, while interacting with the N terminus of the first MCM (Miller et al. 2019).

Cdt1 is crucial for efficient establishment of pre-RCs. However, conflicting data exists on whether it is required for MCM2-7 association with ORC in yeast extracts (Chen et al. 2007; Takara & Bell 2011; Frigola et al. 2013). In vitro data indicates that, although Ctd1 contributes to the stability of MCM2-7 association with ORC, it isn't necessary for the initial recruitment (Frigola et al. 2013). Combined with the fact that Cdt1 is released upon successful loading of MCM2-7 loading onto origins, at least in vitro, it may be possible that one Cdt1 molecule can act as a catalyst to load multiple MCM2-7 hexamers.

3.2 Activation

Once at least one MCM double hexamer has been loaded, the origin of replication has been licensed for initiation in S phase. The process of activating an origin involves numerous factors, some of which interact transiently with the double hexamer and others that join it to form larger complexes. When finalized, the process results in the formation of the active so-called CMG helicase, which is made up of Cdc45, GINS, and MCM2-7 (Moyer et al. 2006). The MCM

helicase is a poor helicase by itself and is only properly activated when in the CMG complex (Ilves et al. 2010)

In order for the active helicase to form, the MCM2 and MCM4 subunits of the MCM helicase must first be phosphorylated by DDK at their unstructured C-terminal tails (Sheu & Stillman 2006; Francis et al. 2009). This phosphorylation facilitates the recruitment of Cdc45 and Sld3 to the double hexamers, followed by association of the non-essential but Sld3-stabilizing protein Sld7 (Kamimura et al. 2001; Heller et al. 2011). Following DDK activity, Sld2 and Sld3 are phosphorylated by CDK in the kinase complex's only known essential role in the origin activation cascade (Tanaka et al. 2007; Zegerman & Diffley 2007). Phosphorylated Sld2 and Sld3 are then able to bind Dpb11, forming a complex that is capable of recruiting GINS and DNA polymerase ε (Muramatsu et al. 2010).

In order for the CMG to become an active replisome, the MCM helicases must switch from encircling double-stranded DNA (dsDNA) to single-stranded DNA (ssDNA). The switch from dsDNA to ssDNA occurs after CMG formation, but before ssDNA-dependent association of replication protein A (RPA) and subsequent DNA polymerase α (Pol α) recruitment (Heller et al. 2011; van Deursen et al. 2012; Watase et al. 2012). The essential protein Mcm10 has been shown to be necessary for this switch through the use of a not-yet-characterized ssDNA gate (Wasserman et al. 2019). In a Mcm10-dependent manner, CMG

helicases melt DNA and extrude the lagging strand to switch from encircling dsDNA to encircling ssDNA (Langston & O'Donnell 2019). The CMG, in complex with Mcm10, Pol α, and in the presence of RPA is then capable of performing bi-directional replication (Yeeles et al. 2015).

3.3 Elongation

DNA is synthesized in a 5'→ 3' direction. As a result, synthesis occurs continuously on the leading strand, upon which the helicase travels, and discontinuously on the lagging strand through the use of constant priming and Okazaki fragment formation. On both the leading and lagging strands, the process starts with the synthesis of a small RNA primer by the primase subunit of Pol α, which is then extended by a few deoxyribonucleotides by the same enzyme complex (Pellegrini 2012). This primer is then capable of being extended by two different polymerases: Pol ε for continuous synthesis on the leading strand and Pol δ on the lagging strand (Pursell et al. 2007; Nick McElhinny et al. 2008). Because of the different polarity of synthesis on the lagging strand, Pol δ only synthesizes about a 165 base-pair fragment, called an Okazaki fragment, before coming upon the 5' end of a previously-synthesized, RNA-primed fragment (Smith & Whitehouse 2012). With its strand displacement ability, Pol δ is able to create a flap, which is then, digested by the flap endonuclease Rad27 (Fen1 in higher eukaryotes) or Dna2, depending on the size of the flap (Bae et al. 2001; Ayyagari et al. 2003).

Both Pol ε and Pol α are known to make contacts with CMG. These contacts are direct for Pol ε and largely indirect and through its interaction with Ctf4 for Pol α, though recent structural evidence points also to some direct contacts of Pol α with the CMG (Gambus et al. 2009; Langston et al. 2014; Georgescu et al. 2015). In addition, both enzymes benefit from their interaction with proliferating cell nuclear antigen (PCNA), which is loaded onto primer-template junctions by replication factor C (RFC) and increases the processivity of Pol α and Pol ε by 100-fold and 2-fold, respectively (Chilkova et al. 2007; Georgescu et al. 2014). Though not directly associated with the replisome, topoisomerase II (Topo II), responsible for removing supercoils generated by advancing replisomes, has also been found to be necessary for replication of fragments larger than ~600 bp in vitro (Yeeles et al. 2015).

3.4 Termination

The molecular details of replication termination have only recently started to be elucidated in eukaryotes. Part of the difficulty in studying termination lies in the fact that, although highly dynamic and involving the clash of numerous proteins, the process is not sequence specific and thus difficult to monitor. Genome-wide studies have found that the termination of replication is generally dictated by the timing of replication initiation of specific origins, with termination events occurring halfway between two origins if they fire at the same time in S phase (Hawkins et al. 2013; McGuffee et al. 2013). Notable exceptions to this observation exist,

including the ends of chromosomes where replication forks do not come head-to-head with other replisomes. In addition, the ~150 rDNA repeats in yeast contain site-specific replication barriers facilitated by the protein Fob1, which stalls replication forks coming from one direction in order to avoid replication-transcription collisions (Kobayashi & Horiuchi 1996).

Replication termination can be thought of as involving two steps: end of replication and removal of replisomes. First, the final tracts of DNA between two converging replisomes must be replicated. In yeast, topoisomerase activity to remove supercoils has been shown to be essential for this function, with some redundancy between Topo I and Topo II (Baxter & Diffley 2008; Fachinetti et al. 2010). In addition to topoisomerases, two helicases Rrm3 and Pif1 have been implicated in being important for efficient fork convergence during termination, with Rrm3 acting specifically at the rDNA loci and Pif1 acting globally (Ivessa et al. 2000; Deegan et al. 2019). Lastly, in addition to its role in replication initiation and lagging strand synthesis, DNA polymerase δ seems implicated in leading strand synthesis specifically at sites of replication termination (Zhou et al. 2019).

Second, the replisomes must be disassembled and removed from DNA. Recent evidence points to ubiquitination of Mcm7 after the completion of replication by the E3 ubiquitin ligase SCF in complex with the F-box protein Dia2. The ubiquitinated Mcm7 (and by virtue of their interdependent association, CMG)

then becomes a substrate for removal by the Cdc48 segregase (Maric et al. 2014).

3.5 DNA Replication: Only once per cell cycle

The replication of eukaryotic genomes only once per cell cycle is crucial for the maintenance of genome stability. Yeast cells accomplish this feat through multiple mechanisms involving differences in kinase activity, with G1 being exemplified by overall low kinase activity and S phase by high kinase activity. The end result of this tight and redundant regulation is the clear separation of two key steps in replication: origin licensing and origin activation. To prevent origin re-licensing, both G1 cyclins (Clns) and S phase cyclins (Clbs) are involved in the phosphorylation of the MCM2-7 helicase, prompting its eviction from the nucleus (Labib et al. 1999). In addition, high kinase activity also phosphorylates Cdc6, targeting it for degradation, as well as ORC, preventing it from performing helicase loading (Drury et al. 2000; Nguyen et al. 2001). Importantly, all three of these mechanisms must be inhibited in order for re-replication to occur in budding yeast (Nguyen et al. 2001).

4. Anatomy of Origins of Replication

4.1 Origins of replication in *S. cerevisiae*

One of the biggest reasons that budding yeast has been such an important organism in elucidating the molecular details of replication is due to the fact that

its origins of replication are sequence specific. Work on yeast origins started with the discovery of autonomously replicating sequences (ARS), chunks of the yeast genome that contained an origin and were able to propagate episomes, starting with the 850 base-pair ARS1 element (Stinchcomb et al. 1979). Since then, many of the mechanistically-relevant parts of the ARS1 and other origins have been elucidated. It is now known that ORC binds an origin at the ARS consensus sequence (ACS), which is made up of an 7 base-pair A/T-rich sequence that is necessary but not sufficient for origin function (Figure 1.2) (Broach et al. 1983; Marahrens & Stillman 1992; Eaton et al. 2010). Indeed, of the many thousands of possible ACS sequence matches, only 300-400 are bound by ORC in vivo (Xu et al. 2006; Eaton et al. 2010). Chromatin footprinting of ORC indicates that the complex binds asymmetrically to the ACS and extends for ~50 base pairs in the 3' direction relative to the T-rich strand (Bell & Stillman 1992). Genome-wide analysis of ORC binding sites has made possible the determination of a consensus sequence for the ACS as well as an extended 32 bp motif defined by ORC-ChIP data that more accurately captures the footprint of

ORC at origins (Xu et al. 2006; Nieduszynski et al. 2007; Eaton et al. 2010).

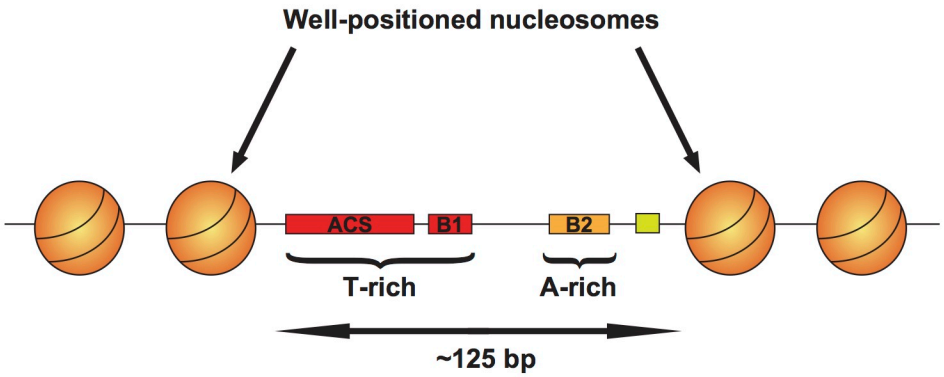


Figure 1.2: Anatomy of yeast origins of replication.
Origins of replication in budding yeast are characterized by a conserved T-rich ARS Consensus Sequence (ACS), which along with B1 constitutes the ORC binding site. The B2 element is found in close proximity to the ACS-B1 elements. It varies in size and composition but tends to be A-rich. Some origins have additional elements next to the B2 element (indicated in lime green) that play important roles in nucleosome positioning and origin efficiency. Lastly, replication origins in budding yeast are conserved by well positioned +1 and -1 nucelosomes. (See text for details and references.)

Although the ACS is crucial for ORC binding and proper origin function in vivo, other DNA elements play a role as well. The most well-characterized origin, ARS1, contains B1, B2, and B3 elements to the 3' side of the T-rich strand of the ACS (Marahrens & Stillman 1992). The B1 element is enveloped by the ORC footprints that have been determined both in vitro and in vivo (Bell & Stillman 1992; Diffley & Cocker 1992). Like the ACS, the B elements are also A/T-rich, with the B2 element specifically being a 9/11 match for the ACS core. As discussed previously, the presence of a B2 element that has some affinity for ORC and resembles an inverted ACS may have implications for the dual-ORC model for loading of MCM double hexamers (Wilmes & Bell 2002; Coster & Diffley 2017). Importantly, mutations in any of the three B elements only have the effect of reducing origin function, whereas their complete abrogation completely eliminates origin function (Marahrens & Stillman 1992). Mutation of the B2

element, for example, reduces loading of MCM at ARS1 (Zou & Stillman 2000). Lastly, the B3 element in ARS1 is known to bind the protein Abf1, whose role seems to involve the proper positioning of nucleosomes close to the origin, which has been shown to be important for origin function (Lipford & Bell 2001; Eaton et al. 2010). It is important to note that, even though the B elements of ARS1 have been characterized to great detail, the same has not been done for most origins. In fact, genome wide analysis of origins indicates that there is high variability in the kinds, numbers, and positioning of B2 and B3 elements (Chang et al. 2011). The B1 element, on the other hand, is better conserved throughout the genome as part of the ORC binding site (Eaton et al. 2010).

Some aspects of origins, however, are applicable very broadly. For examples, functional ACSes are known to reside in nucleosome-free regions (NFRs). Mutations to the ACS that cause nucleosomes to encroach in the NFR and artificial encroaching of ARSes by nucleosomes causes a reduction in their function (Simpson 1990; Lipford & Bell 2001). Notably, altering the nucleosome pattern at ARS1 and even moving well-positioned nucleosomes away from the origin also decreases origin function, indicating that nucleosomes can play both a positive and negative role for ARS1 function (Lipford & Bell 2001).

4.2 Origins of replication in other organisms

Although the molecular details may differ, the underlying laws that govern replication in all domains are strikingly similar. In the bacteria E. coli, for example,

replication initiation is also sequence specific. Initiation in this organism starts at oriC, which is made up of 11 closely-located 9-bp recognition boxes for the initiator protein DnaA (Kaguni 2011). However, unlike yeast and bacteria, higher eukaryotes do not seem to show any sequence specificity for ORC binding and origin establishment. Unlike in budding yeast, plasmid replication in both Xenopus extracts and human cells does not require and ACS-like element and starts at random sequences, with some preference for AT-rich islands in Xenopus (Krysan & Calos 1991; Mahbubani et al. 1992; Stanojcic et al. 2008). Similarly, replication initiation in the fission yeast S. pombe does not have sequence specificity but also prefers asymmetrically AT-rich locations (Segurado et al. 2003). Analysis of ORC binding sites in the fly Drosophila also was not able to find sequence specificity, but was able to predict ORC binding sites based on adjacent sequences that signified its local chromatin environment (MacAlpine et al. 2010). One element that has been proposed to mark sites of replication initiation in mouse, fly, and human cells is GC-rich tracts that are able to form G4 quadruplexes, being present in 80-90% of initiation sites, as well as CpG islands (Cayrou et al. 2011; Besnard et al. 2012). However, these elements are very abundant and not sufficient to confer origin activity in by themselves.

5. MCM Dynamics Throughout the Cell Cycle

5.1 Expression

As for many genes, transcript levels for MCM2-7 in budding yeast cycle throughout the cell cycle, with a peak at M/G1 and a decrease after the start of S phase (Granovskaia et al. 2010). MCM2-7 protein levels are relatively stable throughout the cell-cycle, with half lives of 1.7 hrs (~1 cell cycle) for Mcm7 and 5.9+ hrs for the rest of the components in the hexamer (Hennessy et al. 1990; Christiano et al. 2014). Since the complex is stable, control of its function is largely based on its localization, as mentioned in the previous section.

Accurate quantification of the absolute number of MCM molecules per cell has been challenging and has produced numbers that vary more than an order of magnitude between components of the hexamer and between the same component in different studies. For example, Mcm3 was estimated to be present at 18,000 molecules/cell, 200,000 molecules/cell, and 35,100 in three different studies using western blot techniques or GFP fluorescence (Ghaemmaghami et al. 2003) (Donovan et al. 1997) (Lei et al. 1996). A recent study looked at the reported amounts of protein in 21 separate publications that used different measurement criteria (ex. a.u, mol/cell, molarity, mass) (Ho et al. 2018). Using in-silico methods, the data from all the studies was converted to molecules per cell and the reported median value estimates ranged from 2774-5360 molecules per cell for each of the components of the MCM helicase.

5.2 Localization and stability

The localization of the MCM2-7 helicase throughout the cell cycle in budding yeast is dynamic and tightly regulated. The behavior of each component of the hexamer throughout the cell cycle is fairly uniform. The complex is transported as a hexamer into the nucleus in late mitosis/G1 (Hennessy et al. 1990; Yan et al. 1993; Lei et al. 1996). Degradation of any of the six components of the hexamer inhibits the nuclear accumulation of the rest. In addition, degradation of any of the components in G1 when the hexamer is nuclear leads to the nuclear export of the rest of the components (Labib et al. 2001). Furthermore, nuclear accumulation of MCM2-7 seems to be interdependent on Cdt1, as degradation of Cdt1 prevents MCM2-7 nuclear accumulation and vice versa (Tanaka & Diffley 2002).

After the start of S phase, the MCM hexamer is phosphorylated by B-type cyclin CDKs and excluded from the nucleus to remain cytoplasmic during G2/M (Labib et al. 1999; Nguyen et al. 2000). Mcm2 and Mcm3 both contain weak nuclear localization signal (NLS) peptide, which alone are not sufficient to produce robust nuclear localization but together form a potent NLS (Liku et al. 2005). Importantly, Mcm3 also contains a nuclear export signal (NES) peptide. Phosphorylation of the Mcm3 region that contains the localization modules by CDKs promotes the nuclear export of free MCM2-7 at the beginning of S phase.

Unlike in humans and fission yeast, budding yeast Cdt1 is a stable protein whose function is controlled by dynamic localization (Tanaka & Diffley 2002). Similar to MCM2-7, Cdt1 becomes nuclear in late M/G1 and then is exported out of the nucleus after the start of S phase. Cdt1 is nuclear in an α-factor-induced G1 arrest but becomes cytoplasmic in a Cdc7-ts-induced late G1 arrest. Since Cdc7 is an essential component of the DDK kinase, this result indicates that Cdt1 is exported from the nucleus before replication initiation. MCM2-7, on the other hand, is located in the nucleus in both α-factor and Cdc7-ts arrests (Tanaka & Diffley 2002).

The MCM hexamer can be formed in vitro using separately-purified MCM components, as assayed by gel filtration (Davey et al. 2003). Cdt1 is thought to play its role in MCM2-7 localization not through the contribution of an NLS or generation of the complex but through the stabilization of the hexamer, which in turn allows the previously mentioned NLS-NES module to perform its function (Wu et al. 2012).

5.3 How many MCM double hexamers are loaded at each origin?

Although many advances have been made in determining the molecular mechanism of origin licensing, a question remains as to how many MCM double hexamers are loaded at each origin. In vitro experiments have shown that loaded MCM can slide on double-stranded DNA, raising the possibility that loaded MCMs can freely slide away from their ORC-adjacent loading site to allow more

MCMs to load (Remus et al. 2009). Additionally, single-molecule experiments indicate that ORC leaves its binding site after successful loading of a double hexamer, opening the possibility for another ORC to bind the ACS and another hexamer to be loaded without any prejudice of previous loading events (Ticau et al. 2015). Consistent with this model, the ability of ORC to hydrolyze ATP is necessary for reiterative loading of MCM (Bowers et al. 2004).

Numerous studies in different organisms have presented conflicting data on whether multiple MCMs are actually loaded at origins. Genome-wide yeast studies using chromatin immunoprecipitation (ChIP) and sequencing suggest that only one MCM double hexamer is loaded per origin (Eaton et al. 2010; Belsky et al. 2015). Similarly, single molecule in vitro experiments using lambda DNA chromosomes also only found mostly 2, and at maximum 3, MCM2-7 hexamers bound to origins (Duzdevich et al. 2015).

In separate yeast studies, in vitro single-molecule and electron microscopy studies of specific origins show that more than one MCM double hexamer can in fact be loaded, though at a much lower incidence than loading of only one double hexamer (Remus et al. 2009; Ticau et al. 2015). Bulk experiments using yeast extracts also show that at least two double hexamers are loaded at ARS1 (Bowers et al. 2004). Similarly, another study using quantitative western blots found that ARS1 had on average 3 MCM double hexamers loaded in G1 (Das et al. 2015). Lastly, measurements of in vivo chromatin binding indicate that MCM

binds origins at a much higher number than accounted by the number of origins alone (Donovan et al. 1997).

The loading of multiple MCM double hexamers per origin of replication seems to be more prevalent in higher eukaryotes. Studies in Xenopus egg extracts have shown that multiple MCMs are loaded onto replicons, with estimates of 5 double hexamers per replicon or more than 5 double hexamers per 1kb of DNA (Mahbubani et al. 1997; Edwards et al. 2002). Similarly, studies in Drosophila indicate that multiple MCMs are loaded at origins of replication and tend to distribute away from the ORC binding sites (Powell et al. 2015). Lastly, studies in HeLa cells indicate that there are significantly more chromatin-bound MCM molecules than there are replicons (Burkhart et al. 1995).

5.4 What are the physiological benefits for MCMs to be loaded in excess?

Advances in the understanding of the molecular mechanism of replication fork movement have shown that replisomes only need one MCM helicase during regular replication. These findings would indicate that a double MCM hexamer loaded at each origin should be sufficient for successful replication. However, data showing that more MCM are loaded onto chromatin than are used for replication, and that these MCM often do not co-localize with replication forks, has pointed to the existence of the “MCM paradox.” Though the reason for the excess loading of MCM hasn’t been clarified for every organism, data from some organisms hints at its importance. In Xenopus egg extracts, excess MCM loading

becomes essential for successful replication under conditions of replication stress, indicating that dormant, excess MCM molecules are activated to salvage replicated DNA as a result of the replication stress (Woodward et al. 2006). In *C. elegans,* worms only become susceptible to low doses of the ribonucleotide reductase and replication inhibitor drug hydroxyurea (HU) when their MCM levels are reduced (Woodward et al. 2006). Similarly, studies using human U2OS and HeLa cells have suggested a similar mechanism where excess MCM hexamers that are loaded and not used during unperturbed replication becomes vital for successful replication after treatment with replication stress inducing agents (Ge et al. 2007; Ibarra et al. 2008). Together, these data indicate that excess loading of MCM may play an important role in dealing with replication stress.

An important distinction to note is that excess MCM loading is thought to occur in two contexts. First, MCMs are loaded in excess of the number of origins that normally fire in S phase. These excess MCMs are loaded at dormant origins that can be activated during replication stress, as stated above. Second, MCMs have also been shown to load in excess at single origins of replication and in close proximity to each other. This method of loading has not specifically been implicated in relieving replication stress. Given the ability of MCMs to slide on dsDNA, they could potentially be pushed ahead by active helicases and rescue collapsed forks when needed. However, data supporting such a model has yet to be presented. Importantly, the cellular pool of available MCMs likely determines how and if MCMs are loaded in excess in either context. Therefore, exploring the

consequences of modulating MCM levels has important implications for both mechanisms.

5.6 Consequences of reducing or increasing endogenous MCM levels

Part of the MCM paradox relates to the apparent lack of consequences from the reduction of MCM levels during unperturbed replication. Questions remain as to what are the cellular consequences of either increasing or decreasing MCM levels. In yeast, reduction of Mcm2 and Mcm3 levels in diploids by 50% has no effect on the rate of growth (Lei et al. 1996). However, in the same study, reduction of Mcm2 but not Mcm3 leads to loss of stability of minichromosomes being propagated by ARS1 and ARS404 origins of replication but not ARSH2B2 (aka ARS ARS428), indicating that reduction of MCMs may affect specific origins in different ways. MCM reduction also has significant consequences for fission yeast, with temperature-sensitive mutants of MCM components going through S phase slower and exhibiting reduced viability (Liang et al. 1999).

Studies in mice have also explored the consequences of MCM reduction. Comparison of mouse embryonic fibroblasts (MEF) expressing ~1/3 of normal Mcm2 levels to WT revealed that, under conditions of replication stress, Mcm2-deficient cells are unable to use dormant origins (Kunnev et al. 2010). Importantly, even without replication stress, the survival of mice harboring the MCM-deficient genotype was significantly reduced compared to wild type and they were more prone to tumorigenesis (Kunnev et al. 2010). Analysis of

replication initiation in MCM-deficient MEFs indicated specific sites in the genome had reduced replication initiation compared to wild type (Kunnev et al. 2015). These sites included gene-rich, early-replicating regions, and often correlated with regions exhibiting copy number variations (CNVs) found in tumors from mice harboring the same genotype (Kunnev et al. 2015). However, similar to the yeast studies, the molecular mechanism that leads to the observed phenotypes in both cases is unclear.

Conversely, increase of MCM levels from their endogenous levels has also shown to have adverse effects. For example, specific overexpression of Mcm7 in epidermal tissue of mice leads to significantly advanced tumor formation when these mice are challenged with carcinogens (Honeycutt et al. 2006). Together, these data suggest that deregulated MCM levels have significant implications for proper cell function.

5.7 Levels of MCM and cancer in humans

Due to their essential role in DNA replication and proliferation, MCM levels have been found to be crucial prognostic and diagnostic markers for cancer. Many cancers, including ovarian, lung, breast, thyroid, cervical, prostrate, colorectal, and numerous hematological cancers show increased MCM levels (Giaginis et al. 2010). In some cases, studies have shown that MCM levels are more sensitive indicators of proliferation than conventional markers such as Ki-67 (Kato et al. 2003; Schrader et al. 2005; Chatrath et al. 2006). In addition to its

use as a diagnostic marker, MCM levels have also proven to be strong prognostic markers. Mcm2 levels, for example, are a strong prognostic marker for breast cancer, with high Mcm2 levels indicating lower overall survival rates (Gonzalez et al. 2003).

6. Replication Timing

6.1 What is replication timing and how does it arise?

Eukaryotic cells establish a stringent separation between loading the complexes necessary for replication and their activation. However, the control over a successful round of replication does not end there. Indeed, at the population scale, cells replicate their genome in a temporally ordered manner. Which molecular mechanisms underlie this temporal order, how they establish it, and what the consequences are when this replication timing goes awry is still a matter of intensive research.

Several concepts and metrics have been used over the years in order to fully describe replication timing. First, the concept of replication timing depends on the observation that some origins replicate early in S phase (early origins) while others replicate late in S phase (late origins) (Raghuraman et al. 2001). In any given S phase, each origin has an intrinsic probability of firing that is based on many factors, some of which are understood and others that are actively being investigated. The efficiency of an origin, on the other hand, is defined by the

percentage of cells in a given S phase that activate that specific origin (Rhind & Gilbert 2013). Lastly, every origin and locus in the genome can be described by the timing of replication (t rep), the time at which 50% of the cells in the population have replicated the origin or locus (Raghuraman et al. 2001).

Though the genome-wide replication timing profiles of populations of cells have shown impressive reproducibility, it has become increasingly apparent that these profiles arise from stochastic processes. Using single molecule assays, numerous studies have shown that not every early origin fires every cell and every cell cycle, and not every passively-replicated late origin remains dormant every cell cycle (Czajkowsky et al. 2008). However, averaging the activity of origins from multiple cells still gives rise to the replication timing profiles observed in population studies (Czajkowsky et al. 2008). In addition, mathematical modeling of replication using origin specific data from multiple studies has shown that the timing profiles can be recapitulated using a stochastic model (de Moura et al. 2010; Yang et al. 2010). Nonetheless, even though the stochasticity of replication has been demonstrated, what gives rise to it and how it is implemented during replication remains a subject of investigation. Some of the mechanisms exerting control on specific origins are discussed below.

6.2 Why is replication timing important?

It has become increasingly clear that a defined replication timing program is an evolutionarily conserved characteristic for most eukaryotes (Rhind & Gilbert

2013). In yeast, origins in a given cell population have characteristic firing times during S phase (Raghuraman et al. 2001). In metazoan cells, replication initiates at largely random sites during early embryogenesis when cells are rapidly diving and are transcriptionally silent (Hyrien & Mechali 1993). As development progresses, specific lineages develop specific replication timing, with as much as 50% of the genome changing its timing profiles (Hiratani et al. 2010). The mechanisms guiding these changes aren't clear, but it seems that they are relevant in many disease states. Some cancers, for example, exhibit distinct changes to their replication timing profiles, to the point that these can be used as diseased state signatures (Ryba et al. 2012; Sasaki et al. 2017).

One mechanism by which replication timing can have significant effects is through gene dosage. Genes in regions where replication occurs earlier in S phase may have twice the expression levels of later replicating regions during S phase. Eukaryotic cell have general mechanisms to compensate for gene dosage during S phase (Padovan-Merhar et al. 2015; Voichek et al. 2016). However, these mechanisms do not seem to be universally followed. In yeast, for example, histone genes replicate early in S phase. This characteristic has functional consequences, as delaying the time of replication at histone genes by modifying nearby origins significantly reduces their expression levels during S phase, thus demonstrating a direct cause and effect between replication timing and expression levels (Müller & Nieduszynski 2017).

The timing of replication at specific regions has also been linked to mutation rates. In yeast, late replicating regions can have as much as a six-fold higher mutation rate compared to early replicating regions due to a mechanism likely involving error-free versus error-prone damage DNA repair responses at different times during S phase (Lang & Murray 2011). Similarly, mutation rates in different regions of human cells are also significantly affected by the time in S phase at which they replicate (Chen et al. 2010). The mechanism behind the replication timing establishment in these cells and the consequences for the differences in mutation rates isn't clear, but it may point to replication timing having a major effect on the natural evolution of specific parts of each of these genomes.

Whether changes in replication timing are a cause or consequence of genome instability is unclear. Therefore, gaining a better understanding of the molecular mechanisms that underlie the establishment of replication timing profiles is crucial to solving this puzzle.

6.3 Factors contributing to replication timing in *S. cerevisiae*

6.3.1 Origin architecture

The architecture of yeast origins has been shown to be important for origin function. Nucleosome positioning around the ACS, as established by ORC and specific DNA binding factors such as Abf1, is important for origin function (Lipford

& Bell 2001; Eaton et al. 2010). Although all origins require an NFR to function, there seems to be a correlation between the size of that NFR and the timing of origin firing, with early origins generally having wider NFRs and late origins generally having narrower NFRs (Soriano et al. 2014). In addition, there is a correlation between nucleosome exchange rates and origin firing, with higher exchange rates near origins correlating with earlier firing origins (Dion et al. 2007).

6.3.2 Trans-acting factors and epigenetic modifiers

Numerous trans-acting factors have been implicated in the control of specific origins and genomic regions in well-elucidated mechanisms. Telomeric origins, for example, have been shown to be late-replicating due in part to Rif1 binding and subsequent recruitment of the protein phosphatase PP1, which counteracts the origin-activating kinase activity of DDK and delays origin firing at these loci (Lian et al. 2011; Davé et al. 2014; Hiraga et al. 2014; Mattarocci et al. 2014). The late replicating program at telomeres seems to supersede specific origin regulation mechanisms since early origins that are transferred to telomeric regions become late firing (Ferguson & Fangman 1992). Another example involves the Forkhead transcription factors Fkh1 and Fkh2, which have been shown to be involved in the early firing of specific origins that are proximal to Fkh binding sites by clustering these origins together in the nucleus (Knott et al. 2012). Lastly, the Ctf19 component of the kinetochore complex has been shown

to be involved in the early firing of centromeric origins by recruiting the DDK kinase to these loci (Natsume et al. 2013).

Acetylation of histones is another mechanism that has been implicated in the regulation of origin timing. Tethering of the histone acetyltransferase Gcn5 close to ARS1412, for example, causes this usually late origin to fire early during replication (Vogelauer et al. 2002). On the other hand, the histone deacetylate (HDAC) Sir2 has been shown to be involved in the early firing of some origins while also being responsible for the late firing of other origins, including the rDNA origins (Yoshida et al. 2014). The rDNA-specific role of Sir2 is counteracted by another HDAC, Rpd3, which itself delays origin firing at other groups of origins (Vogelauer et al. 2002). These data point to histone acetylation as being implicated in the regulation of both early and late firing origins.

6.3.3 Location in the nucleus during S phase

An origin's location in the nucleus during S phase has also been shown to be important for when it fires during S phase. Telomeres, for example, are located in the nuclear periphery, as are many late replicating regions during G1 when origins are licensed (Heun et al. 2001). Switching the location of early origin ARS1, which is normally located near the center of the nucleus during G1, with the late origin ARS501, which is located in a telomere, causes the timing of the two origins to be reversed (Ferguson & Fangman 1992). However, this mechanism has been shown to only work for certain instances, or perhaps in

smaller scales, as global disruption of the protein responsible for telomere localization to the nuclear periphery does not make telomeric origins fire early in S phase (Hiraga et al. 2006).

6.3.4 Abundance of initiation factors

Studies have shown that, although both early and late origins are fully licensed for replication in G1, their ability to compete for initiation factors that are in limiting concentrations may be reason for the discrepancy in their initiation times. Indeed, overexpression of the initiation factors Sld2, Sld3, Dpb11, and the DDK kinase components makes late origins fire earlier in S phase (Mantiero et al. 2011; Tanaka et al. 2011). The factors and mechanisms that determine an origin's ability to successfully compete for these initiation factors (some of which were mentioned in previous sections) are actively being investigated.

6.3.5 Differential MCM loading at origins of replication and its effect on replication timing

As detailed in previous sections, excess loading of MCM hexamers in relation to the minimal amount needed for replication and to the overall number of documented origins has been observed in many organisms, including yeast (see section 5.3). Recent genome-wide ChIP work in budding yeast indicates that MCM is loaded at origins in different amounts, and that the level of MCM loading correlates with the time in S phases at which an origin fires (Das et al. 2015).

Early origins tend to have more MCM loaded, while late origins tend to have less MCM. This behavior was also seen for in vivo origins using a quantitative western blot assay, which estimated ARS1 to have about 3 MCM double hexamers loaded in G1 (Das et al. 2015). Importantly, reduction of MCM loading through a mutation in the B2 element of ARS1 caused a delay in origin firing during S phase. Together with other studies showing multiple double hexamers (DHs) loading at origins of replication, these data point to a mechanism where the timing of origin firing, at its most mechanistically basic level, is determined by the number of DHs loaded. Simply put, the more MCMs that are available for activation in S phase, the higher the likelihood that the specific origin will initiate early. Indeed, this model fits well with kinetic modeling of replication, which predicts that stochastic firing arising from differences in the availability of an initiator complex (i.e. MCM) in the presence of limiting initiation factors (Sld2, Sld3, Dpb11, DDK) can give rise to the observed replication kinetics seen in budding yeast cells (Yang et al. 2010).

7. Introduction to Thesis Work

A simple model for genome duplication would suggest that the time at which a specific region of the genome replicates should not matter as long as accurate duplication of all chromosomes is accomplished. Although accurate duplication of the genome is paramount, the specific time during S phase at which loci are replicated has also been shown to be important.

As discussed in previous sections, the levels of the replicative helicase MCM in many organisms, though often in excess of the theoretical amount needed for replication, are tightly regulated and correlate with diseased states when either decreased or increased from their endogenous levels. Importantly, MCM has also been implicated in the establishment of the replication timing program in yeast. In order to test the functional consequences of modulating MCM levels in budding yeast and to further explore its role in the establishment of replication timing, I have performed genome-wide experiments monitoring MCM loading under reduced and overexpressed MCM levels and followed up with surveys of the effects that these changes have on the initiation of replication during S phase.

CHAPTER II: The Effects of Perturbing Cellular MCM Levels on MCM Loading at Origins of Replication and Replication Timing in S phase

1. Introduction

The mechanisms regulating replication timing and the consequences of their misregulation have been a subject of extensive research. Though many factors have been correlated with replication timing, a thorough mechanistic understanding of how the timing programs arise in various organisms has remained elusive.

One model that has been proposed to regulate replication timing at a mechanistic level is by MCM loading (Das & Rhind 2016). Several studies in multiple organisms have indicated that MCM is present in cells in large amounts and loaded on chromatin in excess of what is used for replication (introduced in Chapter I.5). Evidence from budding yeast indicates that origins of replication are able to load multiple double hexamers at origins (Das et al. 2015). In that study, ARS1 was shown to have on average approximately 3 double hexamers of MCM loaded, whereas the late origin ARS316 only had ~0.25, indicating that only about a quarter of the cells in the population had one double hexamer loaded at this origin. In addition, mutation of the B2 element at ARS1, shown to decrease MCM loading in previous studies, led to a reduction of MCM loading as well as a delay in the time of replication of that origin during S phase. Furthermore, the study showed a genome-wide correlation between the MCM signal by ChIP-seq and the origin timing-time parameter n (Yang et al. 2010). n describes an origin's firing time while taking into consideration the chances that it would have been

passively replicated by an adjacent, earlier-firing origin. In certain cases, passive replication of an origin falsely portrays that origin as more efficient and earlier-firing (Yang et al. 2010).

The model of the regulation of replication timing by the number of MCMs loaded at origins agrees well with the observed stochastic origin firing observed in yeast and models of replication timing predicting an initiator complex that is present in varying abundance at different origins (Czajkowsky et al. 2008; Yang et al. 2010). For example, the model predicts that an origin that has three DHs loaded has three times the probability of having one of them activated during S phase compared to an origin that has one DH loaded. Under this model, most origins would have at least one MCM double hexamer loaded. This point is important as it ensures that multiple dormant origins are available to for use if necessary, as has been shown to be crucial for cell viability under conditions of replication stress (Woodward et al. 2006).

In order for the multiple-MCM model to be feasible, origins must accommodate a large footprint of MCM double hexamers. The NFR at origins of replication has an average size of ~125bp and can be as high as ~190bp at some origins (Eaton et al. 2010; Soriano et al. 2014). Considering the ~48bp footprint of ORC-CDC6 and the ~68bp footprint of one MCM double hexamer, for a total of ~113bp, most origins would only be able to load one MCM double hexamer (Speck et al. 2005; Remus et al. 2009). So, how could multiple DHs be

loaded onto origins? As mentioned in the introduction, the size of the nucleosome-free region (NFR) varies in size at origins and a wider NFR correlates with earlier firing origins (Soriano et al. 2014). An origin with a ~190bp NFR, as seen in some cases, could theoretically accommodate two double hexamers (Soriano et al. 2014). In addition, nucleosome exchange rates around NFRs in yeast are known to vary, with some nucleosomes leaving DNA and re-binding at higher rates than others (Dion et al. 2007). Importantly, the exchange rates of nucleosomes adjacent to early origins tend to be higher than the exchange rates at late origins. Combined with the observation that stably-loaded MCM double hexamers can slide on dsDNA, this raises the possibility that some DHs may slide past the binding sites of well-positioned nucleosomes around origins of replication.

There are several questions about this multiple MCM model that can be addressed by perturbing MCM levels. First, does the amount of MCM loading change when the cellular levels of available hexamers are changed? If so, what do these changes tell us about how MCM is loaded at origins of replication? There are at least two ways that the number of MCM loaded at origins could be regulated. In one model, the amount of MCM loaded is determined by the capacity of said origin to accommodate MCMs (i.e. “origin capacity model”), with some origins being able to accommodate more DHs than others (Figure 1.3). This capacity to load DHs can be mediated by available space in the ORC-adjacent area or chromatin factors that can effectively increase that space. This

possibility is supported by studies correlating wider NFRs and higher nucleosome exchange rates with earlier firing origins.

A second method by which the number MCMs loaded at origins of replication could be regulated relies on ORC activity (i.e. “ORC activity model”) (Figure 1.3). In this model, each origin has a specific ORC activity that dictates its ability to load MCM. This activity may be mediated by ORC affinity for that origin sequence as well as by chromatin remodeling and modifying factors. In budding yeast, in vitro ORC affinity for a specific origin sequence correlates well with its affinity in vivo only for a subset of replication origins (Hoggard et al. 2013). These origins are described as “DNA dependent” whereas origins whose affinity for ORC in vivo cannot be explained by its affinity in vitro are described as “chromatin dependent.” Indeed, recent in vitro studies show that chromatin remodeling factors have strong effects on origin licensing, with ISW2 specifically blocking the ability for ORC to bind DNA (Azmi et al. 2017). In additional support of this model, in vivo ORC binding levels have been shown to be correlated with MCM binding levels (Das et al. 2015)

These two models can be tested by perturbing MCM levels in budding yeast. In the origin capacity model, origins would be filled with MCM in a distributive way, as supported by single molecule data showing that ORC is released from DNA after one double hexamer is loaded (Ticau et al. 2015). For example, one can imagine a scenario under unperturbed conditions where ORC binds one

origin and loads a DH. ORC then releases from that origin and binds another one to load another DH, and so on and so forth for multiple origins. As more MCM are loaded, some origins reach their capacity and only the ones that are able to accommodate additional MCM double hexamers do so. In this model, reduction of cellular MCM levels would most affect origins with large amounts of MCM, causing their levels to drop. Increases in the cellular pool of MCM, on the other hand, may increase MCM loading at higher peaks if these are not saturated, or keep the levels unchanged if MCM loading is already saturated under endogenous conditions.

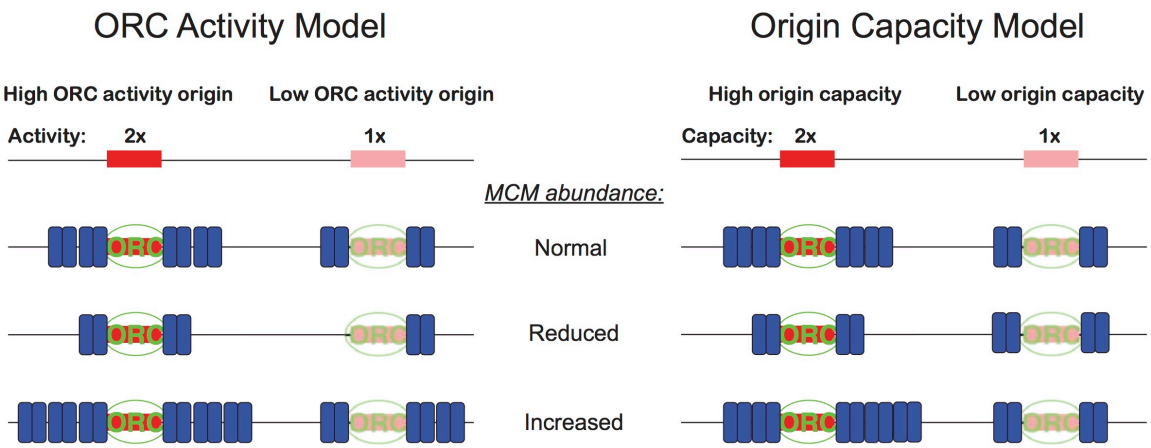


Figure 1.3: Potential models for how multiple loading of MCM double hexamers occurs in budding yeast.

In the ORC activity model, MCM loading at origins of replication would occur differently. Here, an origin with twice the ORC activity compared to another origin would load twice the amount of MCM. In this model, decreases to the cellular pool of MCM levels would be predicted to reduce MCM levels proportionally for all origins. On the other hand, increases to the cellular levels of MCM proteins

would be predicted to cause increases to the levels of MCM at all origins, if these are not already in excess under endogenous conditions.

In addition to the mechanism of how multiple MCMs are loaded at origins of replication, it remains unclear at a genome-wide scale what the consequences are when MCM loading is changed. The multiple MCM model predicts that changes to MCM loading would correlate with changes in replication timing. For example, if the MCM loading at a specific origin were reduced as a consequence of reduced cellular MCM levels, the initiation time of that origin would be delayed. This effect was confirmed when MCM loading at ARS1 was perturbed, but whether it holds true for multiple origins genome wide remains to be tested (Das et al. 2015).

In order to test the effects of altered MCM levels on how MCM is loaded at origins of replication and to see what implications these changes may have for replication timing, I perturbed cellular MCM levels in budding yeast. Using genome wide chromatin immunoprecipitation and replication timing assays, I found that lowering cellular MCM levels had significant effects on MCM loading and replication timing. On the other hand, increasing the MCM levels in budding yeast had no effects on MCM loading onto DNA or replication timing in S phase.

2.1 Increasing MCM levels in yeast to assess chromatin loading at origins of replication and effects on replication timing

2.1.1 Cells overexpressing the full MCM2-7 hexamer are viable

In order to gauge the effect of increased MCM2-7 in cells, we employed the galactose overexpression system in yeast. In this system, genes being driven by galactose-inducible promoters turn on their expression when galactose is added to the media. MCM components were cloned into bidirectional Gal-1,10 promoters and integrated into the genome wild type yFS1020 yeast (Ticau et al. 2015). Additionally, the overexpressed Mcm7 was tagged at its C terminus with GFP in order to help monitor its overexpression levels. In order to test whether cells overexpressing MCM components are viable, cells were spotted by serial dilution on YP-Raffinose (no overexpression) or YP-Galactose (overexpression) plates. Overexpression of single MCM components or the complete hexamer did not affect cell viability (Figure 2.1). Tagging of Mcm7 with GFP also showed no adverse effects on growth. Interestingly, overexpression of the MCM hexamer in conjunction with overexpression of the loading factor Cdt1 was not viable (Figure 2.1). This data shows that MCM2-7 overexpression is viable and can be used for further experiments.

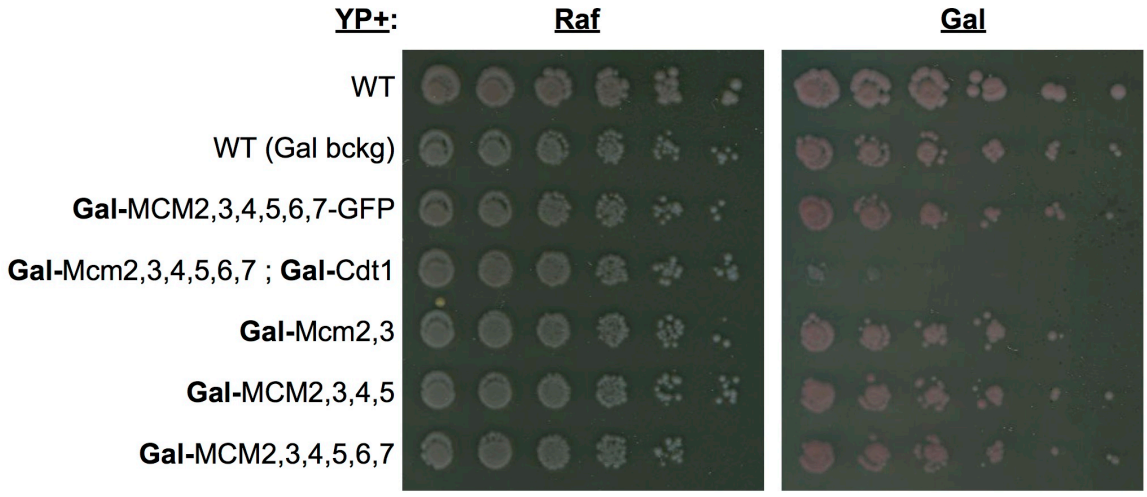
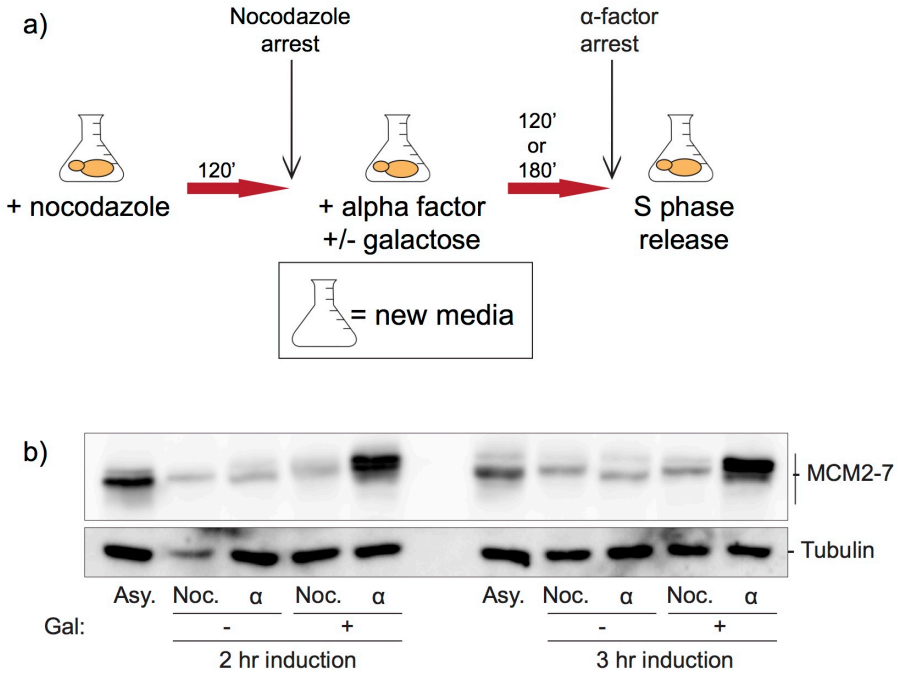


Figure 2.1: MCM overexpression does not affect cell viability. Spot assay of log phase cells of the indicated genotypes.
Cells overexpressing specific proteins are marked by 'Gal-' followed by the overexpressed proteins. 'WT' refers to the Rhind lab WT strain (yFS833). 'WT (Gal bckg)' refers to the non-MCM-overexpressing strain (yFS1020) to which Gal overexpression vectors were added. The rest of the strain names: Gal-MCM2,3,4,5,6,7-GFP (yFS1075); Gal-MCM2,3,4,5,6,7, Gal-Cdt1 (yFS1021); Gal-MCM2,3 (yLD172); Gal-MCM2,3,4,5 (yLD173); Gal-2,3,4,5,6,7 (yFS1076).

2.1.2. Galactose induces robust overexpression of MCM during a G1 arrest and helicase overexpression does not affect S phase progression

We wanted to determine whether overexpression of the MCM2-7 hexamer led to changes in helicase loading during G1, and whether any of these changes had implications for replication timing. To that end, we synchronized cells in G2 using nocodazole then released them into media containing α-factor to synchronize them at the G1/S boundary before DNA replication. In addition to α-factor the media also contained either galactose or raffinose, to induce or repress MCM2-7 overexpression, respectively (Figure 2.2a). The arrest was carried out for two or three hours in order to evaluate whether prolonged G1 arrest leads to differential or increased MCM loading. Induction of MCM2-7 was monitored using a polyclonal antibody against MCM2-7. After two or three hours, robust

overexpression of the hexamer was observed (Figure 2.2b). A portion of the α-factor arrested culture was collected for ChIP-sequencing analysis while the rest was released into S phase to monitor replication timing. As seen in Figure 2.3, overexpression of MCM2-7 during a G1 arrest does not have any significant consequences for cell cycle progression. A slight delay in initiation for the cultures that were induced with galactose is attributed to the change in sugar sources.



c)

Fold-increase in MCM2-7 signal

		2 hr induction	*3 hr induction*
- gal	**Noc**	1.0	1.0
	Alpha	1.6	1.0
+gal	**Noc**	1.0	1.0
	Alpha	5.3	5.2

Figure 2.2: Galactose robustly induces overexpression of MCM2-7.
a) Experiment outline. b) Western blot gauging level of overexpression of MCM2-7 upon galactose induction for 2 or 3 hours. c) Quantification of MCM2-7 signal in α-factor arrested cells relative to levels during a nocodazole arrest.

2.1.3 Genome-wide distribution of MCM2-7 in overexpressing cells corresponds to known origins of replication

In order to measure MCM loading in cells overexpressing the full hexamer, we performed micrococcal nuclease (MNase) digestion followed by chromatin immunoprecipitation of MCM using a polyclonal antibody against the full hexamer (Wal & Pugh 2012). Fragments of the genome that were protected by MCM were pulled down, sequenced, and their abundance was normalized to an S. pombe spike-in control. Figure 2.4a shows that this method was successful in identifying previously annotated origins of replication, with peaks corresponding to most of the annotated origins in the budding yeast origin of replication database OriDB (Nieduszynski et al. 2007). In addition, the density of MCM reads at specific origins also correlated well with similar experiments from our lab (Figure 2.4b) as well as previous publications using the same antibody (Figure 2.4c). Lastly, MCM levels at origins showed correlation with the timing factor n, as previously published (Figure 2.4d) (Yang et al. 2010). Overall, these data indicate that the MNase ChiP-seq results presented here are reliable measurements for the location and abundance of MCM in the genome.

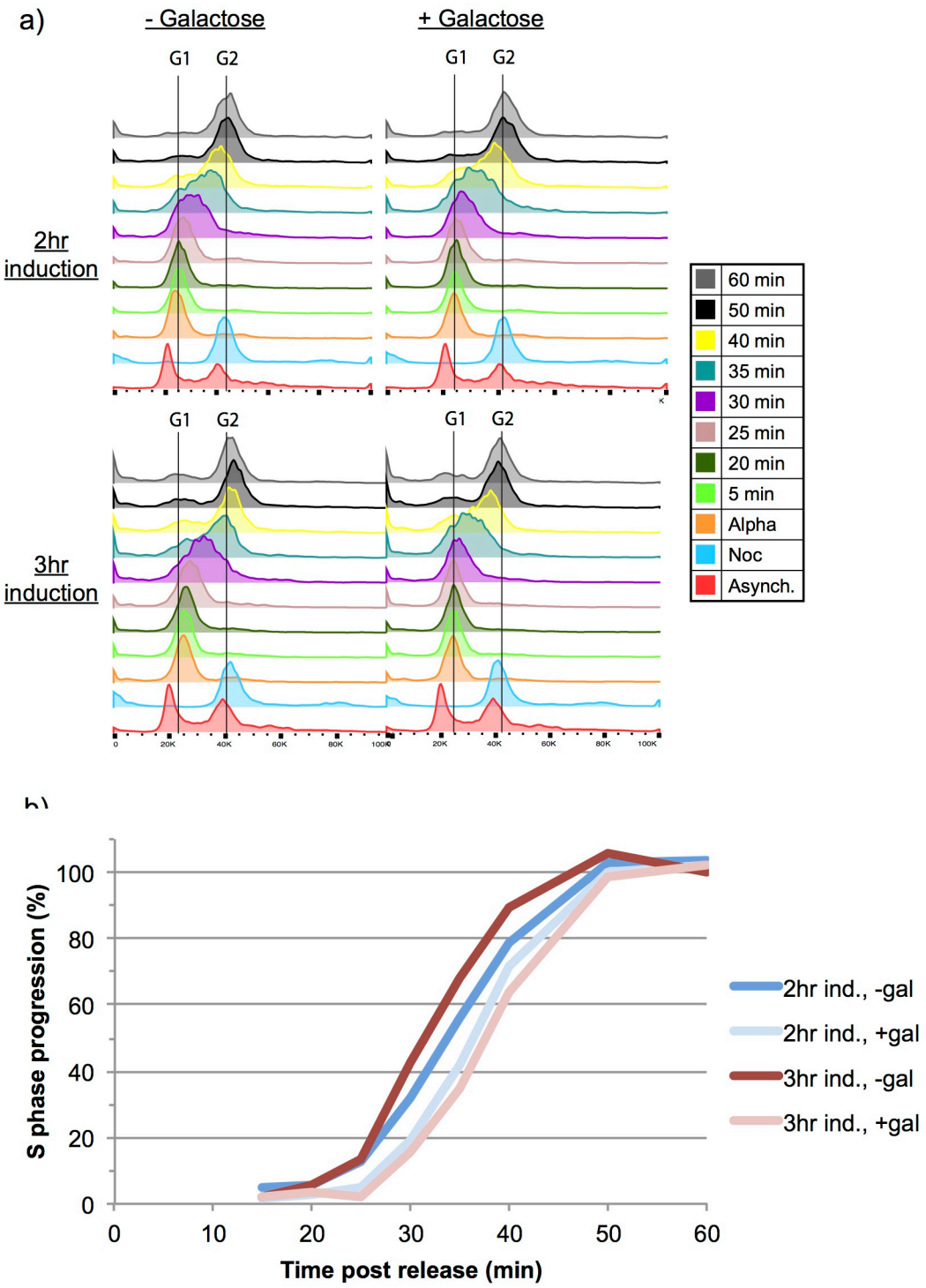


Figure 2.3: MCM2-7 overexpression does not affect progression through S phase.
a) Flow cytometry profiles for cultures released into S phase as outlined in Figure 2.2a. b) Quantitation of the replicating population of cells as they progress from a G1 arrest to a G2 DNA content.

2.1.4 Overexpression of MCM2-7 does not alter levels of helicase loading at origins

To gauge whether increased cellular pools of MCM helicase led to changes in loading during G1, we compared MCM abundance at origins of replication in cells that overexpressed MCM2-7 and those that did not. As seen in Figure 2.5a, there was no significant difference in MCM abundance at characterized confirmed ARS origins when cells were arrested for two hours or three hours in α-factor. Importantly, comparison of MCM abundance at origins from cells overexpressing MCM to those with endogenous levels did not show any differences in loading for either a two hour (Figure 2.5b) or three hour (Figure 2.5c) α-factor arrest. Together, these data indicate that overexpression of the MCM2-7 helicase does not cause altered loading dynamics in G1.

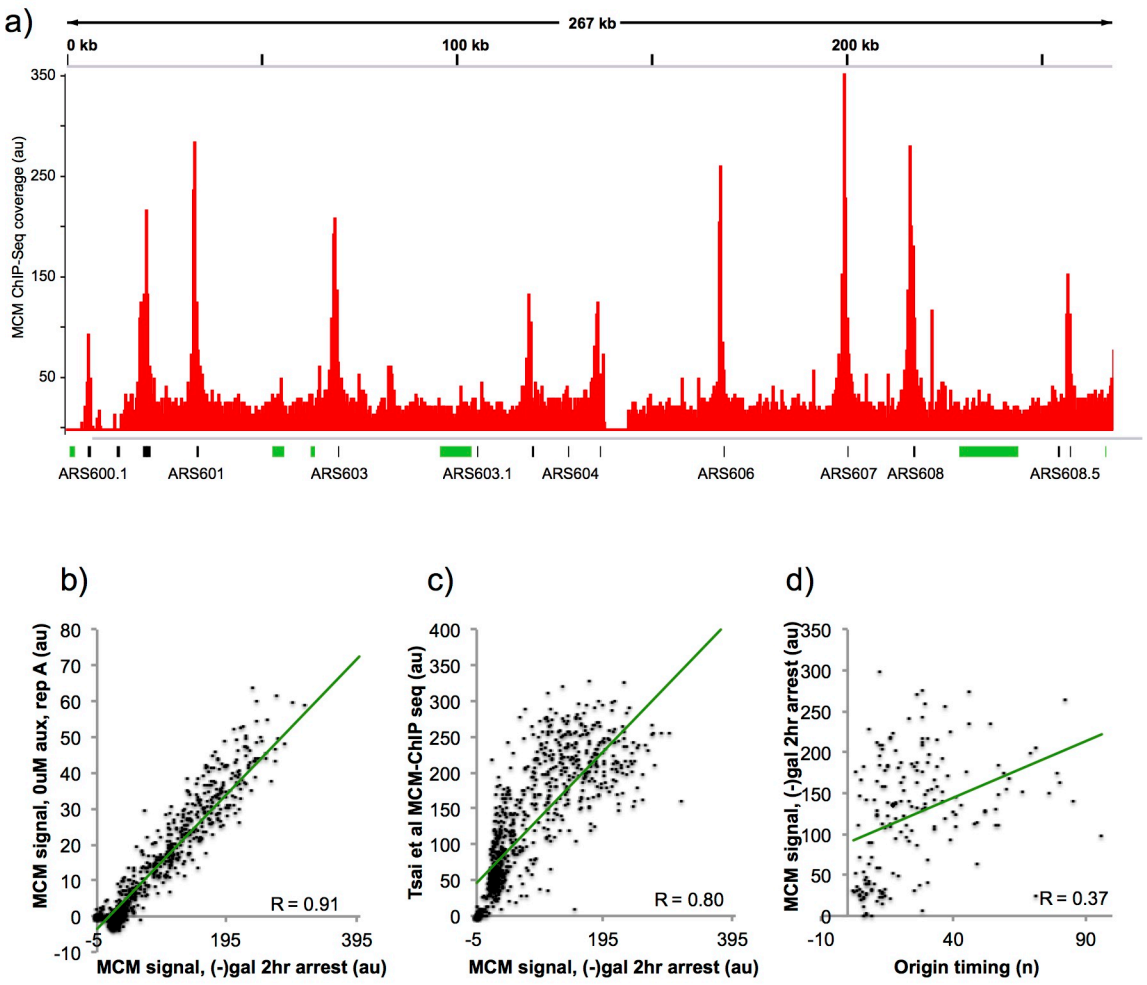


Figure 2.4: ChIP-seq results are reproducible and correspond to known origins of replication.
a) ChIP-seq coverage on Chromosome III for the 2hr arrest in alpha factor, (-) galactose experiment is shown along with with annotated origins of replication. Black lines - ARS origins, green lines - Null origins (not confirmed). ChIP-seq results correlate well with (b) 0uM auxin data and with (c) Tsai et al, (2015) data for all origins on OriDB. d) MCM-ChIP signal at Das et al origins correlates with the timing n.

2.1.5 Overexpression of MCM2-7 does not affect replication timing

In order to relate any changes in MCM abundance to changes in the replication timing program, we performed sync-seq experiments (Müller et al. 2014). Briefly, these experiments are able to monitor replication timing by measuring genome-

wide copy number at specific points after synchronous release into S phase. From this data, the parameter of T rep (time at which 50% of cells have replicated a specific locus) can be extracted. As seen in Figure 2.6a, T rep values correlate well with values obtained from experiments using a separate strain (auxin experiments, later sections). However, comparison of origin replication times for cells overexpressing MCM2-7 to those that were not shows that increased cellular pools of helicase do not significantly alter replication timing (Figure 2.6.b). Altogether, this data suggests that overexpression of the MCM helicase does not alter replication dynamics, possibly due to levels already being in excess and additional reasons further discussed in sections to follow.

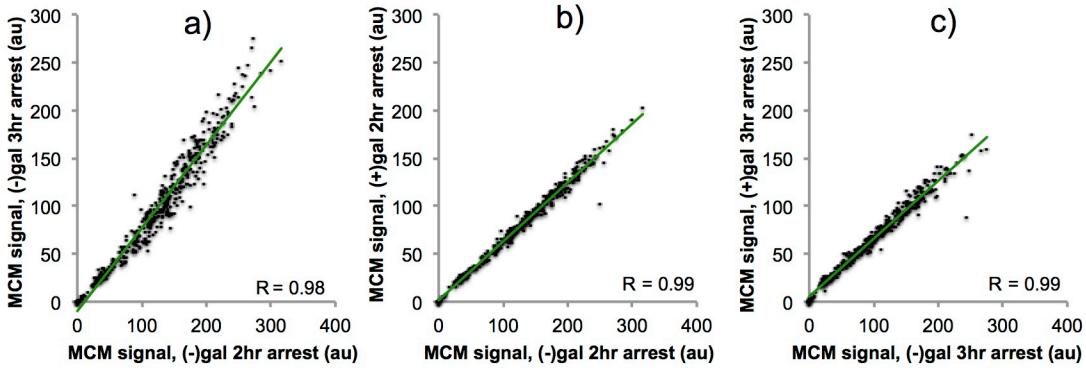


Figure 2.5: Figure 2.5: MCM2-7 overexpression does not cause changes in helicase loading.
MCM signal at ARS origins from the indicated treatments. a) Alpha factor arrested cells at two hours compared to three hrs show no difference in loading. Overexpression of MCM2-7 through galactose-inducible promoters does not change MCM loading after (b) two or (c) three hours in alpha factor.

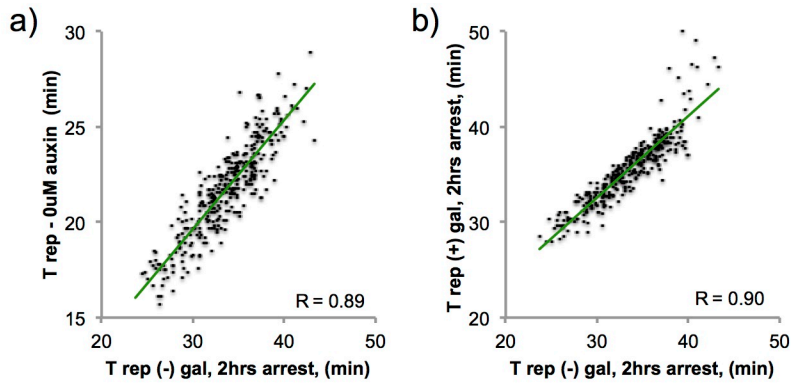


Figure 2.6: Overexpression of MCM2-7 does not alter replication timing.
a) Replication timing in the galactose inducible strain is very similar to replication timing in the auxin inducible degron strain. b) Overexpression of MCM2-7 does not cause any significant changes to replication timing.

2.2 Reducing MCM levels in yeast to assess chromatin loading at origins of replication and effects on replication timing

2.2.1 Auxin-induced degradation of Mcm4 causes reduced viability and sensitivity to hydroxyurea

In order to test the effects of reduced MCM pools on helicase loading and replication timing we employed the auxin-inducible degron (AID) system optimized for yeast (Nishimura et al. 2009). In this system, a protein of interest (POI) is tagged with the degron cassette IAA17. Following addition of the small molecule auxin to media, the degron-tagged POI is targeted for degradation via an interaction with the F box protein TIR1, which is part of the degron strain background. Importantly, degradation levels can be tuned based on the amount of auxin that is added to the media (Nishimura & Kanemaki 2014). In order to reduce the cellular pool of MCM helicase, we tagged one of the six obligate

components of the helicase, Mcm4, with the degron cassette and GFP at its C terminus. As discussed in Chapter I.5.2, the MCM hexamer requires all six components for efficient nuclear localization. In addition, degradation of any single component of the hexamer causes destabilization of the rest (Labib et al. 2001). Therefore, degradation of Mcm4 is expected to reduce the total cellular pool of functional MCM helicases. Serial dilutions of cells harboring the tagged protein showed that it does not affect cell viability, with or without an additional GFP tag (Figure 2.7). However, addition of increasing amounts of auxin caused lethality in cells harboring degron-tagged Mcm4. Furthermore, degradation of Mcm4 made these cells sensitive to the drug hydroxyurea (which depletes dNTP pools and causes replication stress) as seen by the reduced viability in 100mM HU plates compared to wild-type (Figure 2.7). Together, these data indicate that AID is an efficient system to reduce Mcm4 levels in cells and that reduction of these levels leads to reduced viability and increased sensitivity to replication stress.

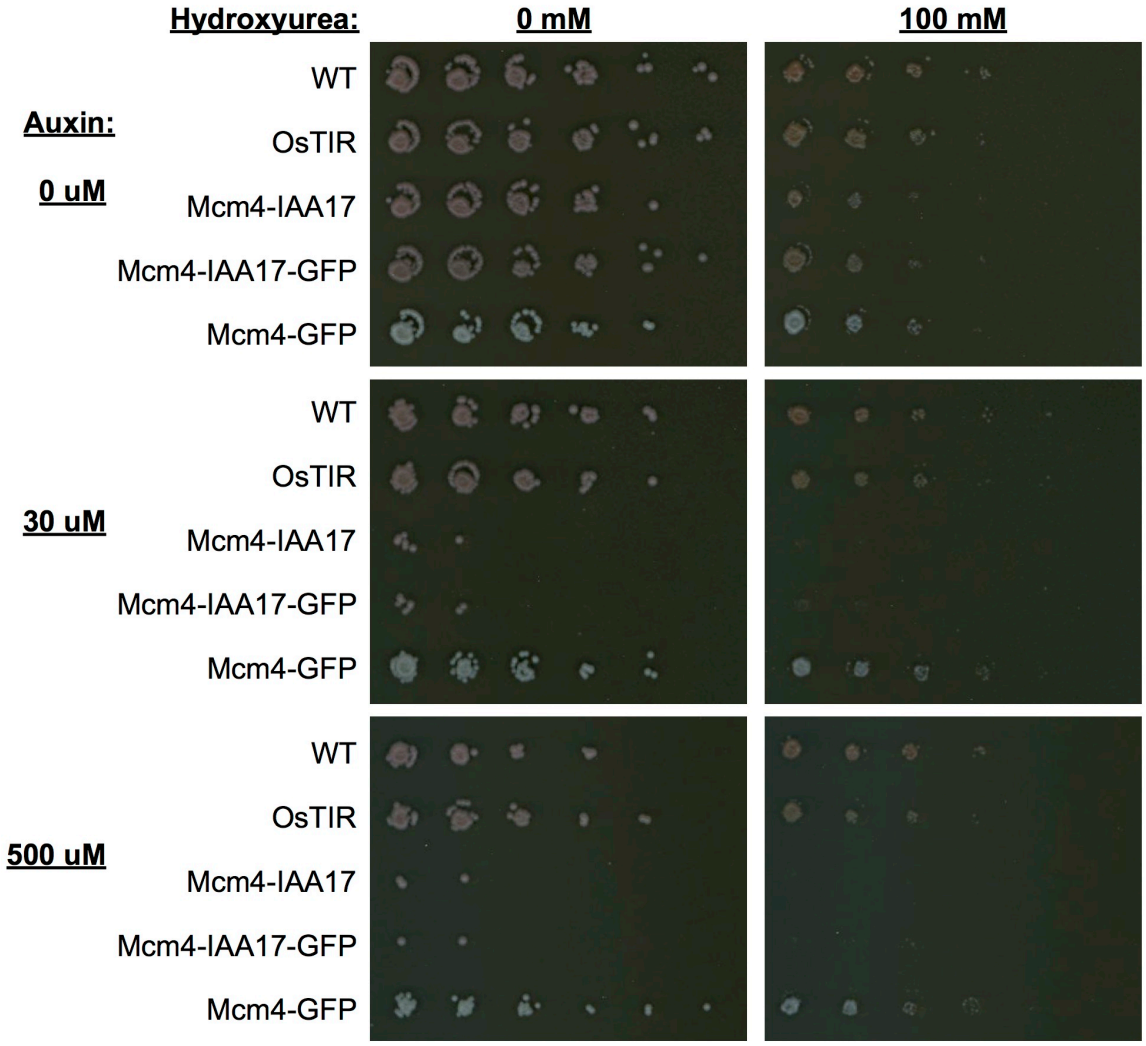
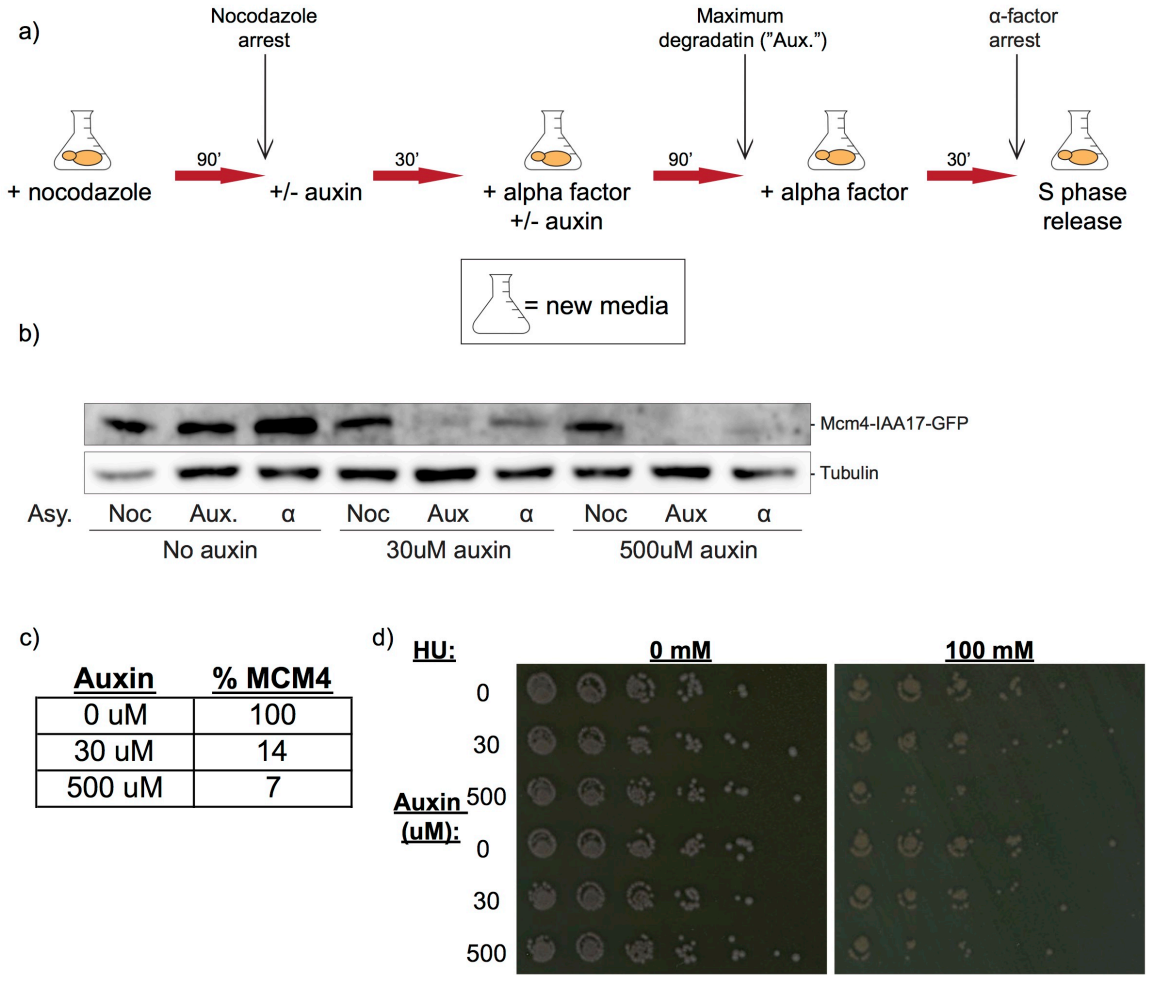


Figure 2.7: Mcm4 degradation causes reduced viability and sensitivity to hydroxyurea. Spot assay of log phase cells of the indicated genotypes, with or without the addition of auxin or HU. Strain names: WT (yFS833); OsTIR (yFS1044), Mcm4-IAA17 (yFS1045); Mcm4-IAA17-GFP (yFS1059).

2.2.2 Auxin-induced degradation of Mcm4 during G1 is dose-dependent and causes slower progression through S phase

MCM is loaded onto origins of replication during G1. To test whether lowered levels of functional helicase affect genome-wide MCM abundance and replication timing, we first synchronized the cells in G2 using nocodazole. Thirty minutes

before release from nocodazole, cultures were treated with 0uM, 30uM, or 500uM auxin to induce Mcm4 degradation. We deemed the synchronization and reduction of Mcm4 levels prior to G1 to be important, as this would ensure that Mcm4 was not removed from active replisomes or pre-loaded origins. Following nocodazole arrest, cells were released into media containing α-factor and the same concentrations of auxin as in the G2 arrest for 1.5 hours. At this point, cells were removed from auxin and allowed to equilibrate in a α-factor-induced G1 arrest for an additional 30 minutes before release into S phase (Figure 2.8a). Protein gels monitoring the levels of GFP-tagged Mcm4-IAA17 revealed that Mcm4 is efficiently degraded after addition of auxin and that the level of degradation was dependent on the dose of auxin that was added (Figure 2.8b). Quantification of Mcm4 levels revealed that, compared to untreated cells, 30uM auxin addition caused cells to have only ~14% of the regular pools of Mcm4, while 500uM auxin further reduced those levels to ~7% (Figure 2.8.c). Importantly, the differences in MCM levels seemed to be consequential for cell viability, as cells released into HU showed a dose-dependent increase in sensitivity compared to untreated cells (Figure 2.8d). Together, these data show that the AID system can be used to reduce MCM levels in a dose-dependent way. Additionally, reduction of Mcm4 pools for only one loading cycle is enough to cause lowered resistance to replication stress.



Auxin	% MCM4
0 uM	100
30 uM	14
500 uM	7

Figure 2.8 : Reduction of Mcm4 pools through auxin-mediated degradation.
a) Experiment outline. b) Western blot of Mcm4-IAA17-GFP levels at the indicated points of the experiment. c) Quantification of Mcm4-IAA17-GFP at α-factor G1/S arrest relative to levels of untreated cells. d) Cultures were synchronized using α-factor as outlined in panel (a), washed, and spotted on YPD plates with or without 100mM HU by serial dilutions.

To test the effects of these reduced Mcm4 levels on S phase progression, we performed flow cytometry analysis of cell cycle progression from cells treated with 0uM, 30uM, and 500uM auxin. As seen in Figure 2.9, degradation of Mcm4 through AID causes a dose-dependent delay in initiation of replication and slower

progression through S phase, indicating that reduced pools of Mcm4 have functional consequences for replication in S phase.

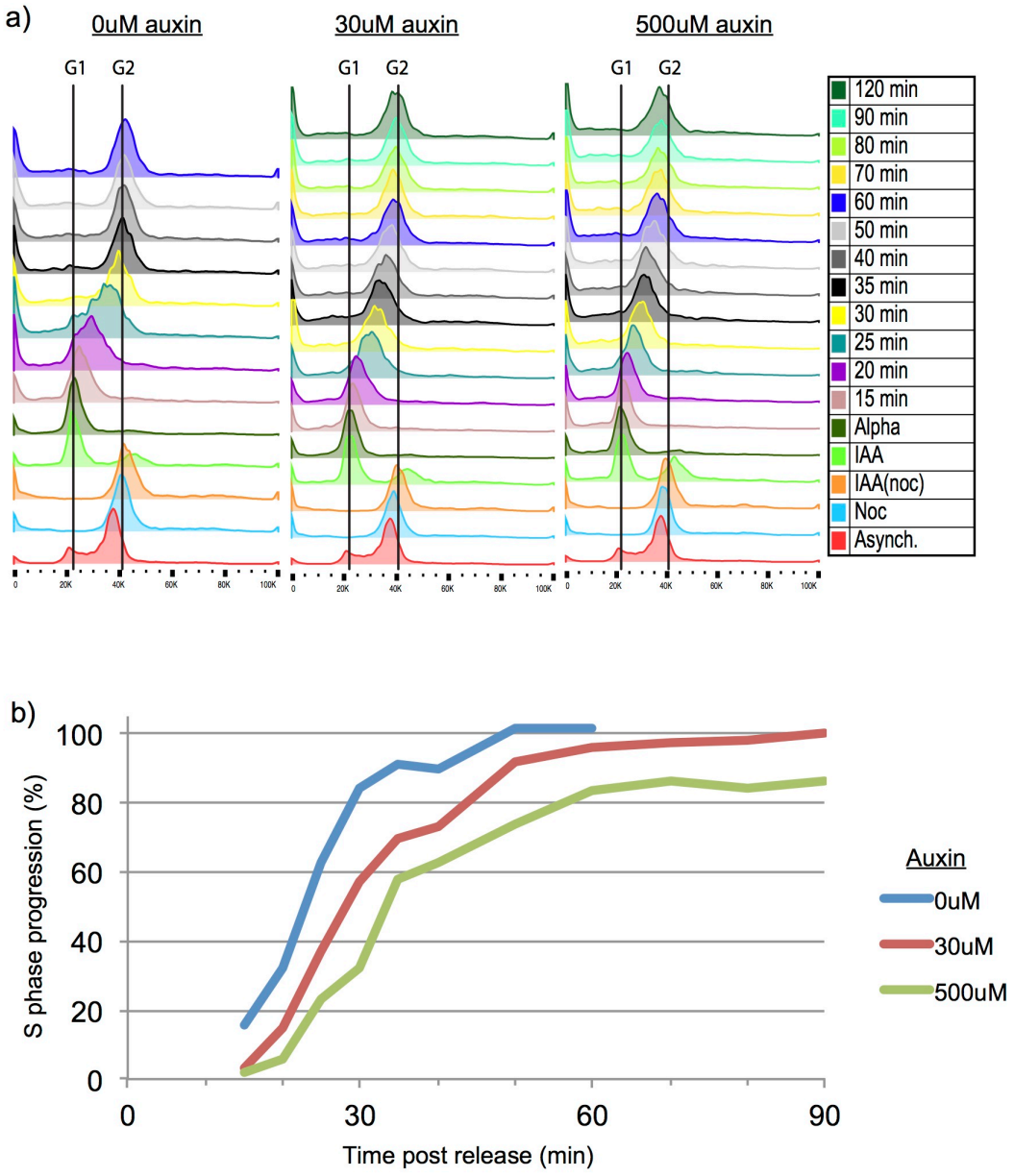


Figure 2.9 : Auxin-mediated degradation of Mcm4 causes slower progression through S phase.
a) Flow cytometry profiles for cultures released into S phase as outlined in Figure 2.8a a. b) Quantitation of the replicating population of cells as they progress from a G1 arrest to a G2 DNA content.

2.2.3 Genome-wide distribution of MCM2-7 in cells with reduced levels of Mcm4

2.2.3.1 MCM2-7 localizes to known origins of replication at varying abundance

Previous ChIP experiments have shown that MCM abundance at origins of replication varies throughout the genome (Chapter I.5.3). To test how lowered MCM pools affect the dynamics of helicase loading and abundance at origins of replication, we performed MNase ChIP-Seq experiments on α-factor-arrested cells after treatment with 0uM, 30uM, or 500uM auxin. As seen in Figure 2.10a, MCM peaks of varying height were successfully localized to known origins of replication. These results were highly reproducible, as a biological replicate of the same experiments showed high correlation (Figure 2.10b). MCM abundance also correlated fairly well with previously published MCM-ChIP results (Figure 2.10c). In addition, MCM abundance for untreated cells correlates with the timing factor n, supporting previous evidence showing that the higher the amount of MCM at an origin, the earlier in S phase it activates (Figure 2.10d)(Yang et al. 2010; Das et al. 2015).

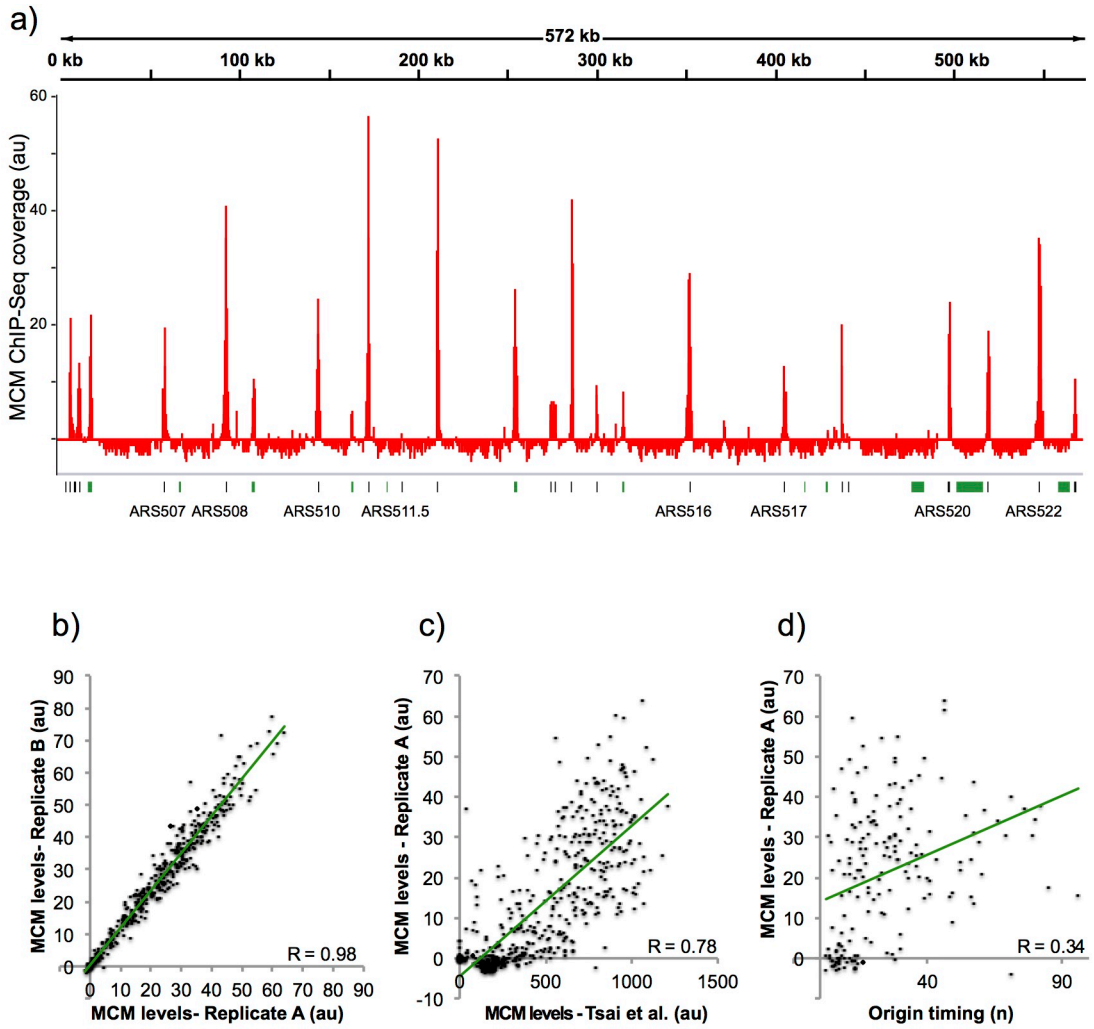


Figure 2.10 : ChIP-seq results are reproducible and correspond to known origins of replication.
a) ChIP-seq coverage on Chromosome 5 for the 0uM Auxin experiment is shown along with with annotated origins of replication. Black lines - ARS origins, green lines - Null origins (not confirmed). ChIP-seq results correlate well between replicates A and B for ARS origin list (b) and with Das et al, (2016) origin list and MCM-ChIP data (c). MCM-ChIP signal at Das et al origins correlates with the timing n.

2.2.3.2 Mcm4 degradation in G1 causes a reduction in helicase loading at origins of replication

Measurements of MCM abundance at ARS origins of replications show changes to MCM loading throughout the genome. As seen in Figure 2.11a, many origins

experience reduction in MCM abundance when treated with auxin, as seen by a move away from the x=y line and towards the x-axis. The level of reduction is dose-dependent, with cells treated with 500uM auxin showing greater reduction in MCM abundance than those treated with 30uM auxin. Plotting for the fraction of MCM abundance that is lost in response to auxin treatment shows that some origins lose most of their MCM signal after 30uM auxin treatment, while others are resistant to the changes in MCM levels (Figure 2.11c or origins colocalizing with x=y line in Figure 2.11b). Treatment with 500uM auxin, on the other hand, causes reductions in MCM levels for most all the ARS origins. Importantly, plotting of the fraction of lost MCM abundance versus MCM abundance in untreated cells shows that low abundance origins are disproportionately affected by reductions in MCM pools (Figure 2.11c).

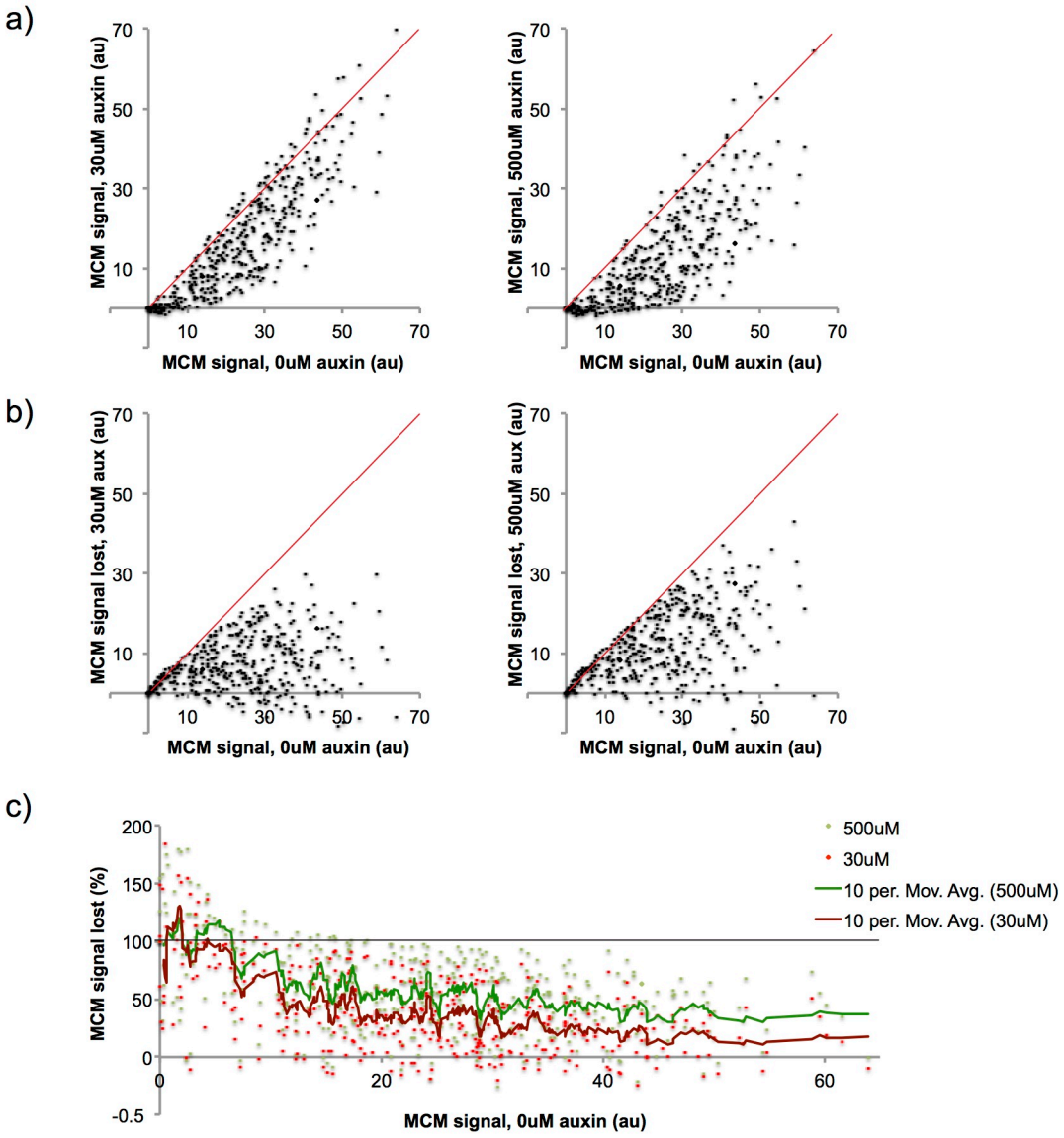


Figure 2.11 : Mcm4 degradation causes a reduction in helicase loading at origins.
a) MCM signal at ARS origins from the indicated treatments. x=y line drawn in red throughout. b) MCM signal for 0uM auxin plotted against the amount of signal lost from auxin treatments. c) Fraction of MCM signal lost upon treatment with 30uM or 500uM auxin. (10 pt moving average fit. y =1 line drawn in black for reference).

2.2.3.3 “Weak” origins are more prone to losing abundance of MCM following reduction of Mcm4 levels

Previous analysis of various replication origins in budding yeast has shown origins can be classified based on their affinity for ORC. Hoggard et al. (2013) measured the ORC affinity of numerous origins in vitro using gel shift assays as well as in vivo using ORC ChIP-seq. Origins whose high in vitro affinity for ORC explained their high in vivo affinity were classified as “DNA-dependent”. Origins whose high in vivo affinity could not be explained by their in vitro affinity were classified as “chromatin-dependent”. Lastly, origins that displayed low affinity for ORC both in vivo and in vitro were classified as “weak”. Figure 2.12 shows the amount of MCM signal lost by these three categories of origins after treatment with 500uM auxin. Interestingly, “DNA-dependent” and “chromatin-dependent” origins are more resistant to changes in MCM pools than “weak” origins. These data indicate that origins with low ORC affinity are outcompeted by origins with high ORC affinity (due to sequence or chromatin reasons) when the pool of available MCM helicases reduced.

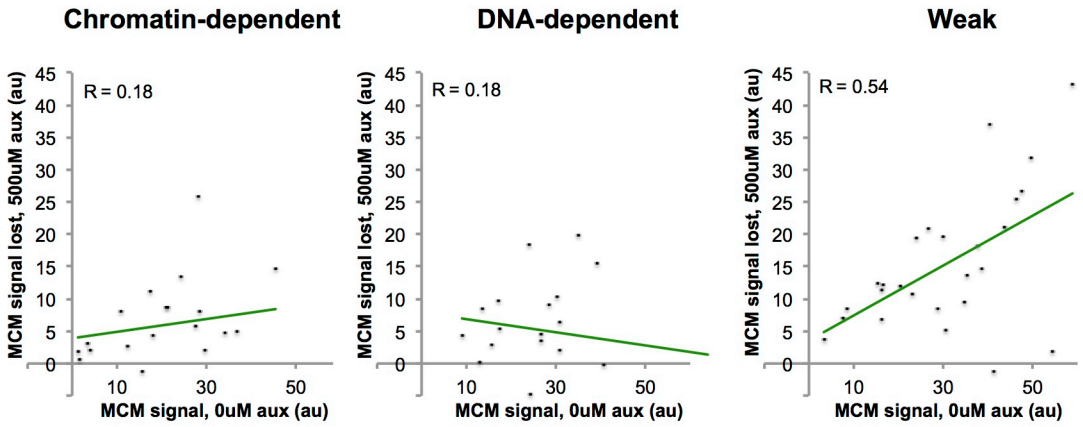


Figure 2.12: Changes in MCM signal at three different categories of origins.
MCM signal at origins of replication characterized as chromatin-dependent, DNA-dependent, and weak in Hoggard et al. (2013).

2.2.3.4 V plots indicate that MCMs associates with flanking nucleosomes and the NFR

Previous studies using MNase footprinting indicate that MCM associates with well-positioned nucleosomes flanking the ACS at origins of replication (Belsky et al. 2015). To gauge the location of MCM in our data, V plots displaying reads around ACS sequences were constructed (Figure 2.13) (Eaton et al. 2010). V plots display the abundance of reads and their length as a function of their genomic location. The abundance and length of reads isolated from chromatin immunoprecipitation of MCM2-7 indicated that MCM associates with 2-3 flanking nucleosomes on either side of the ACS, as seen by the presence of ~60-70bp reads in the ChIP profiles. Contrary to previous reports, MCM signal was also observed in the NFR (Figure 2.13) (Belsky et al. 2015). As expected, the abundance of MCM-sized reads decreased in samples treated with auxin.

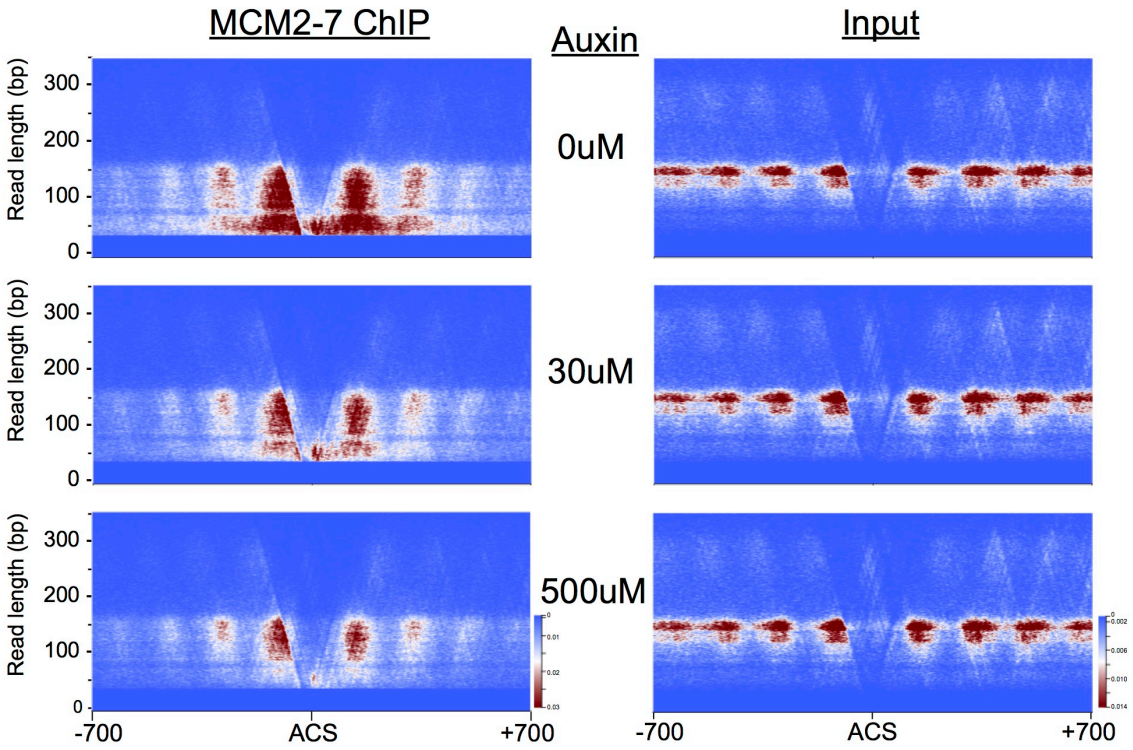


Figure 2.13: MCMs associate with the origin-adjacent nucleosomes and the NFR. MCM ChIP-seq and Input reads were plotted by length on a 1400bp window centered at 253 yeast ACSes (Eaton et al, 2010).

2.2.4 Reduction of Mcm4 levels causes significant changes to the replication timing profile

2.2.4.1 Auxin-induced changes in MCM abundance cause changes in replication timing at specific origins

Reduction of MCM pools via auxin-induced Mcm4 degradation caused significant changes to MCM loading in G1. To test whether the changes in MCM abundance have an effect on replication timing, we performed sync-seq experiments on cells released into S phase after treatment with 0uM, 30uM, or 500uM auxin. Figure 2.14 shows T rep values for Chromosome V as well as the corresponding ChIP-

seq profiles for treated and untreated cells. Overlaying of T rep profiles for this specific locus reveals that some origins that are active in untreated cells become inactive in auxin-treated cells (Figure 2.14b). Interestingly, loss of origin activity correlates with a reduction in MCM signal by ChIP, as is evident for ARS512 and ARS520 (Figures 2.14b and 2.14c). Comparison of T rep values between untreated and auxin-treated conditions revealed and Mcm4 degradation leads to changes in replication timing for a large number of origins and that these changes are larger at higher auxin concentrations (Figure 2.15a). Together, these data demonstrate that reduction of MCM abundance as a result of reduced cellular pools of MCM cause changes to the replication timing profile in yeast.

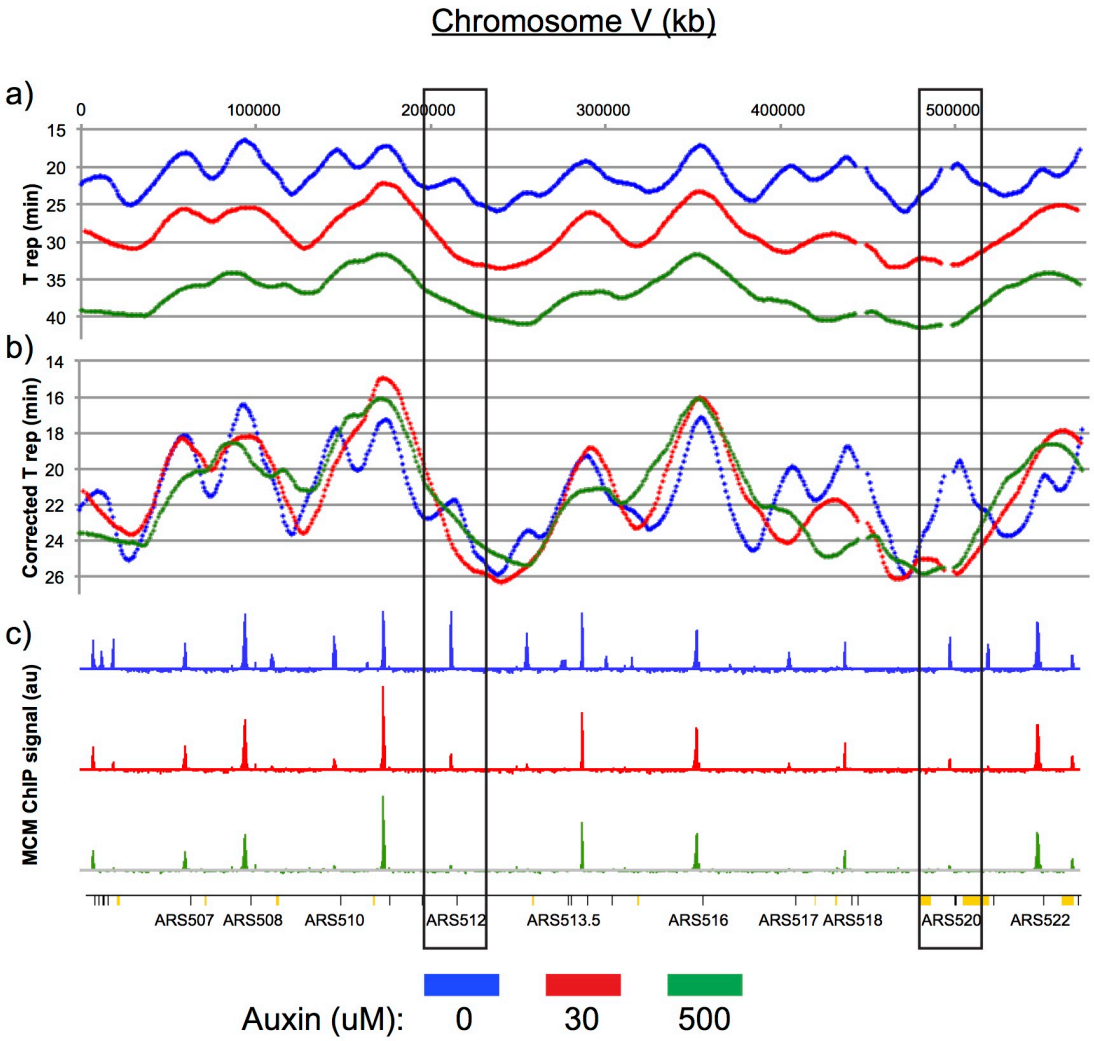


Figure 2.14 : Mcm4 reduction causes changes in replication timing that correlate with MCM2-7 abundance.
a) Chromosome V T-rep values calculated in 1kb windows and LOESS-smoothed. b) T-rep values that have been corrected for a delay in replication initiation. c) MCM2-7 ChIP-Seq data for 0uM, 30uM, and 500uM auxin treatments. Some origins noted. Orange boxes denote "null" origins from OriDB. Boxes denote origins which exhibit significant reductions in MCM signal as well as changes in replication timing.

2.2.4.2 Delays in replication timing at origins of replication correlate with reduction of MCM levels

The replication timing of a number of origins throughout the budding yeast genome is controlled by specific mechanisms involving trans-acting factors (see

Chapter I.6.3). To assess whether changes in MCM abundance at origins of replication correlate with changes in replication timing genome-wide, we measured the T rep values and MCM abundance for a set of origins of replication that excludes those known to be affected by specific mechanisms of origin regulation (Das et al. 2015). Comparison of T rep values and MCM abundance for these origins indicated a small negative correlation between helicase abundance and timing of replication, suggesting that origins with higher MCM levels replicate earlier in S phase (Figure 2.15b). This correlation became stronger when comparing the replication timing of origins in conditions where MCM pools were reduced (Figure 2.15b, 30uM and 500uM auxin). Furthermore, comparing changes in T rep with changes in MCM abundance as a result of auxin treatment revealed a positive correlation, indicating that larger losses of MCM signal correlate with stronger delays in replication timing (Figures 2.15c and 2.15d). Put together, these data suggest that decreases in MCM loading as a result of lower cellular helicase pools lead to delays in replication timing at origins of replication.

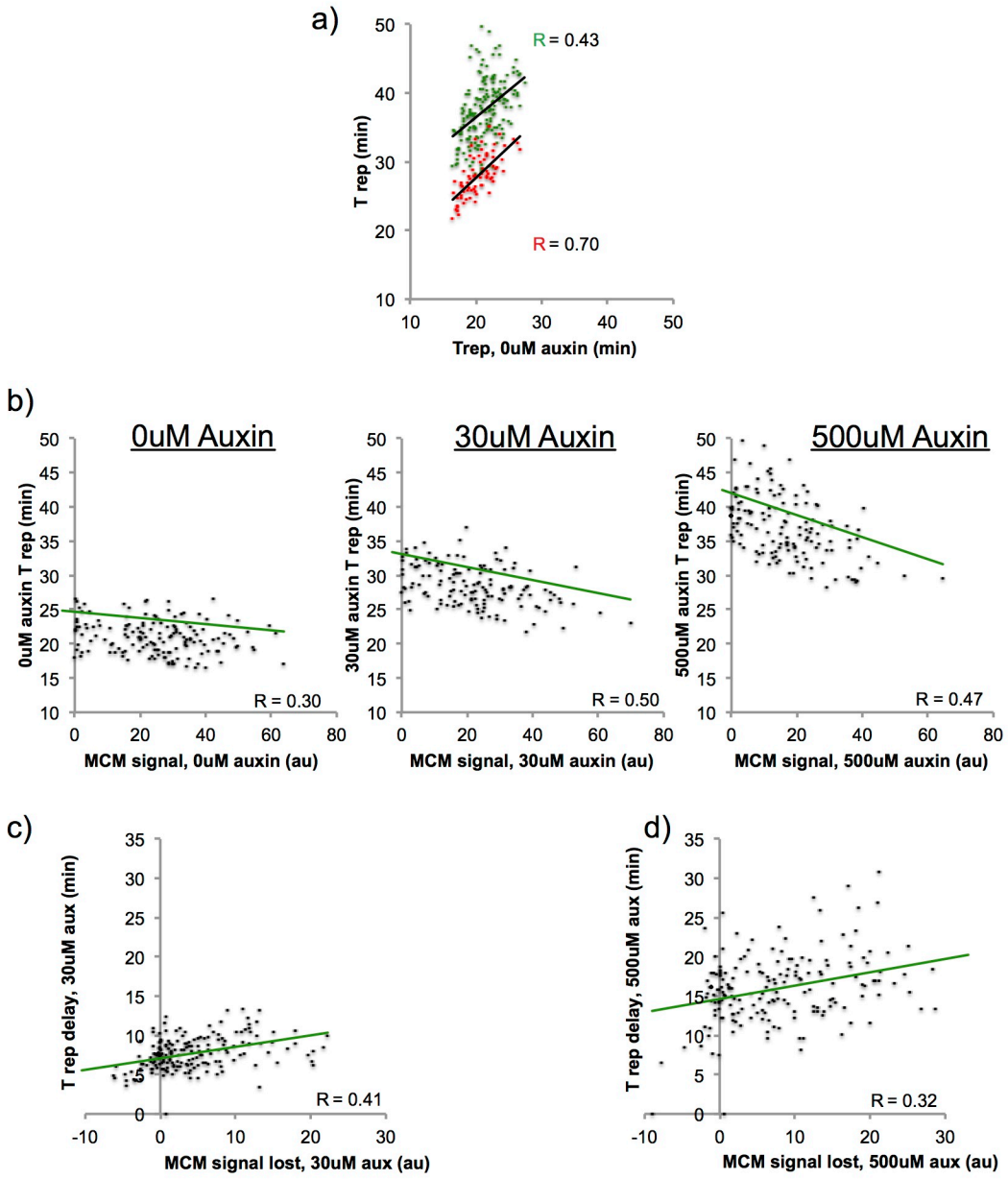


Figure 2.15 : Delays in replication timing correlate with reduction in MCM levels.
T rep and MCM values for Das et al. (2016) origin set are plotted as indicated in the graphs. a) Mcm4 degradation causes genome-wide changes in replication timing. Red=30uM auxin, Green=500uM auxin. b) Delayed replication timing correlates with lower MCM levels for untreated cells. Delays in replication timing correlate with changes in MCM levels for both (c) 30uM auxin treatment and (d) 500uM auxin treatment.

2.3 Discussion

The replication timing program in budding yeast is highly reproducible at a population level. However, the fundamental mechanisms that contribute to the establishment of this program are unclear. Recent evidence suggests a model where loading levels of the main replicative helicase, MCM, may contribute to the establishment of the replication timing program (Das et al. 2015). In this model, origins that load more MCM double hexamers in G1 have a higher likelihood of being activated by limiting factors in S phase. Though multiple MCM loading has been shown in many organisms, including budding yeast, the dynamics of how that loading occurs genome wide and the implications that it has for replication timing remain largely untested.

In order to probe the dynamics of MCM loading during G1 and to directly test its effects on replication timing, we undertook experiments in which the cellular pools of available MCM hexamers were either increased or reduced. Increase of MCM levels through overexpression of the six components of the hexamer did not have any effect on cell viability or cell cycle progression. Consistent with these results, there were no significant changes to MCM abundance at origins of replication or any changes to the replication timing profile of cells released into S phase after MCM overexpression.

In contrast, decrease of functional MCM pools through auxin-induced degradation of Mcm4 had significant implications for cell viability, MCM loading,

and replication timing. Degradation of Mcm4 reduced the viability of cells and made them more sensitive to HU-mediated replication stress. In addition, cells released into S phase after Mcm4 degradation displayed a delay in replication initiation and slower progression through S phase. Genome-wide ChIP-seq of MCM2-7 revealed an auxin dose-dependent reduction in the abundance of helicases loaded at replication origins in G1. Origins that normally load low levels of MCM under endogenous conditions lost a larger fraction of MCM signal as a result of Mcm4 degradation. In addition, origins that show low ORC affinity in vivo and in vitro were more prone to losing MCM abundance after auxin treatment compared to origins that are better able to recruit ORC. Importantly, a reduction in MCM abundance by ChIP correlated with a delay in replication timing for many origins throughout the genome.

The data from our galactose overexpression experiments suggests that increasing MCM2-7 levels has no consequences on MCM loading onto chromatin or replication timing. This conclusion is consistent with data suggesting that MCM is already in excess in budding yeast cells (Donovan et al. 1997). In designing the experiment, we wanted to specifically test the effect of modulating helicase levels. However, previous studies indicate that the MCM2-7 ring requires Cdt1 to enter the nucleus and to be loaded onto origins of replication (Tanaka & Diffley 2002). Although Cdt1 disassociates from the MCM complex after it's successfully loaded onto DNA, it doesn't exit the nucleus until late G1/early S phase. Therefore, while Cdt1 likely shuttles in and out of the nucleus in some equilibrium

during the α-factor arrest in our experiments, it may not be able to accommodate nuclear import for all of the overexpressed MCM2-7 hexamers. As a result, although MCM2-7 is successfully overexpressed in our experimental paradigm, we cannot exclude the possibility that we are only increasing levels in the nucleus to the extent allowed by the nuclear import dynamics of MCM. It's important to note that overexpression of MCM2-7 by itself is not lethal to budding yeast cells but becomes lethal when combined with Cdt1 overexpression (Figure 2.1). Although this points to the synergistic effect of these two complexes, the exact mechanism for the lethality is unclear.

The auxin-induced degradation of Mcm4 caused a delay in replication initiation as well as the speed of progression through S phase (Figure 2.9). The slower progression in S phase is consistent with less MCM being loaded onto some origins and fewer origins being activated in S phase (Figure 2.14b). If fewer origins are activated, it would take longer for the ones that do initiate to replicate the genome. On the other hand, the mechanism behind the delay in replication initiation, as seen by flow cytometry and a global increase in T rep, is unclear (Figure 2.14 and 2.15a). A global and non-specific decrease in MCM loading and origin efficiency as a result of reduced helicase pools may result in delayed initiation. However, several points argue against this possibility. First, replication initiation in budding yeast is limited by a set of initiation factors (Mantiero et al. 2011; Tanaka et al. 2011). If MCM levels were to be lowered genome-wide, the competition for initiation factors would decrease and initiation of the remaining

helicases would in theory be advanced, which is not the case in our experiments. Second, a global decrease in MCM peak heights in our coverage files would be accompanied by an increase in background signal if the data is simply normalized to counts per million (CPM) reads. However, to ensure that the observed peak heights reflect in vivo loading of MCM, we normalized our data to the non-origin “background” signal between all experiments. Therefore, we suspect that the global delay in initiation of replication is not due to a global decrease in origin licensing but instead may be due to a secondary effect of the activation of the degradation machinery, or by a mechanism that we do not understand.

As previously discussed, for the multiple-MCM model to be feasible more than one MCM double hexamer must be accommodated in the relatively tight space of the ACS nucleosome free region. Previous studies have shown that the average size of the NFR is ~125 bp, suggesting that it can only accommodate one DH (Eaton et al. 2010; Soriano et al. 2014). In order for more than one DH to be loaded, the chromatin environment around the origin must then be modified. Previous studies have shown that MCM associates with the two proximal nucleosomes around the ACS (Belsky et al. 2015). Our experiments agree with this data and extend the genome-wide ACS footprint of the MCM hexamer to 2-3 nucleosomes on either side of the ACS, with the likelihood of MCM being found farther from the ACS decreasing with each additional nucleosome (Figure 2.13). In addition, we also found MCM-sized reads inside the NFR. The extended

footprints of MCM in and around the NFR in our data provide support for a model where MCMs are mobile after being loaded and can slide past nucleosomes during nucleosome exchange. Indeed, this data is consistent with observations that early origins, which tend to have more MCM signal by ChIP-seq, display higher rates of nucleosome exchange (Dion et al. 2007). However, considering the fact that sequencing results only provide an average of the footprints of MCM throughout the genome, we cannot exclude the possibility that only one MCM loads at each origin and that the location of that MCM varies from origin to origin.

As discussed in the introduction, there are at least two models for how multiple MCMs can be loaded genome-wide at origins of replication (Figure 1.3). One model relies on the physical capacity of origins to accommodate multiple MCMs. In this model, origins with wider NFRs and higher nucleosome exchange would be expected to accommodate more MCM double hexamers. A decrease in the cellular pool of MCM would affect high MCM origins more than low MCM origins if this model were true. Another model for the loading of multiple MCMs relies on ORC activity at origins of replication. Origins in this model load multiple MCMs due to high ORC activity, perhaps due to the high affinity of ORC for the origin DNA sequence or the presence of specific chromatin factors. A decrease in the cellular pool of MCM would affect all origins proportionally to their activity if this model were true. With these predictions in mind, data from our experiments points to ORC activity as being the dominant mechanism for how multiple MCMs are loaded at origins. MCM levels of origins throughout the MCM loading

spectrum are reduced upon Mcm4 degradation. In addition, we observe that MCM levels of low MCM abundance origins decrease disproportionally more when MCM becomes limiting, suggesting that the high MCM abundance origins are better suited to compete for MCM. In support of this conclusion, analysis of well-characterized origins that are DNA- or chromatin-dependent (and therefore posses high ORC activity) revealed that they are more resistant to a reduction in the available pool of MCM hexamers (Figure 2.12).

It is important to note that the origin capacity and activity models are not mutually exclusive. In fact, as introduced in Chapter I, the size of an NFR as well as the exchange rate of nucleosomes flanking it are correlated with early replicating origins and are likely important general mechanisms for loading at all origins (Dion et al. 2007; Soriano et al. 2014). The origin capacity at specific origins appears to be important for high activity origins, whose MCM levels don't change upon Mcm4 degradation. These origins load a characteristic level of MCM under various levels of MCM pools, indicating that even though their activity is very high, they are still guided by a finite physical capacity around them for the maximal amount of MCM to be loaded (Figure 2.16).

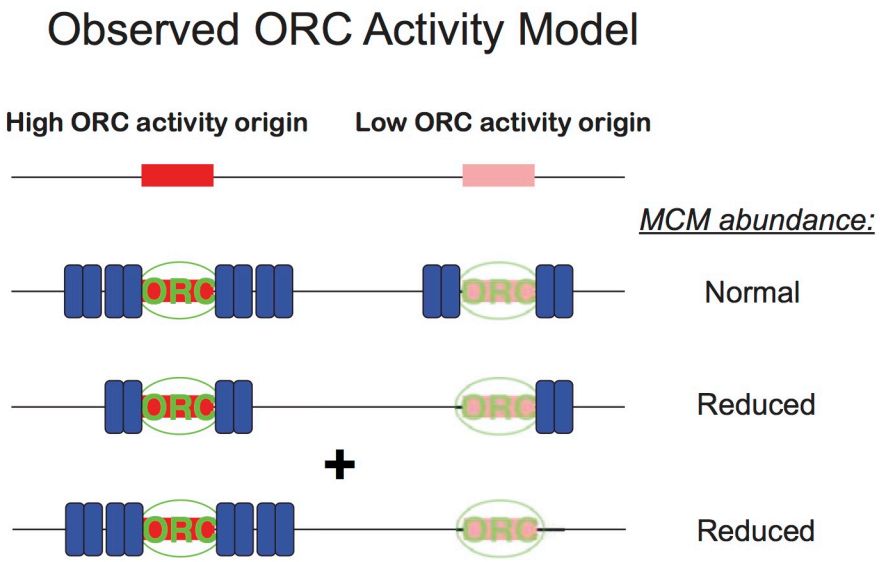


Figure 2.16 : ORC activity is the dominant mechanism of genome-wide MCM distribution. Model for how ORC activity at origins of replication dictates how reductions in MCM levels would affect MCM loading at origins. (See text for details)

Our experimental outline allows direct comparison between MCM abundance at origins of replication and replication timing profiles in the same population of cells. This characteristic makes it possible to test the multiple MCM model at a genome-wide level. In concordance with previous publications, MCM abundance at origins of replication correlates with the timing factor n (Figure 2.10d) (Das et al. 2015). To further test this model, we compared changes in replication timing to changes in MCM abundance that resulted from degradation of Mcm4 and therefore a decrease of the available pool of MCM helicases. This data showed a correlation between loss in MCM abundance and a delay in replication timing. Altogether, these experiments support a model where the level of MCM loading at origins of replication, as directed by ORC activity, has strong implications for when an origin fires in S phase.

Materials and Methods

Strains, Media, and Harvesting

Galactose overexpression experiments

Yeast strain yFS1075 was grown in YP-raffinose at 30°C to an OD(600) of ~0.2. Nocodazole was added to a final concentration of 10ug/ml. After 2 hours, yeast was were vacuum filtered and resuspended in fresh YP-Raf and supplemented with α-factor to a final concentration of 25nM as well as either galactose to a final concentration of 2% or an equal volume of water. The cells were allowed to arrest in α-factor for either 2 or 3 hours before release via vacuum filtration into fresh YP-Dextrose supplemented with 0.2mg/ml pronase (Sigma P6911).

For ChIP samples, α-factor arrested budding yeast cells were crosslinked with formaldehyde and collected as previously described with the following modifications (Wal & Pugh 2012): after formaldehyde quenching, budding yeast cells were combined with log-phase wild-type fission yeast cells (yFS105, also crosslinked) at a 9:1 ratio by OD.

Auxin degradation experiments

Yeast strain yFS1059 was grown in YPD at 30°C to an OD of ~0.2. Nocodazole was added to a final concentration of 7.5ug/ml for 2 hours. 1.5 hrs into the nocodazole arrest, the cultures were supplemented with 30uM final auxin (Indole-

3-acetic acid sodium salt, Sigma item number I5948), 500uM final auxin, or no auxin. After 2 hours in nocodazole, yeast was vacuum filtered and resuspended in fresh YPD supplemented with 5ug/ml alpha factor and 30uM final auxin, 500uM final auxin, or no auxin as before filtration. After 1.5 hours, yeast was vacuum filtered and resuspended in fresh YPD and supplemented with 30uM final auxin, 500uM final auxin, or no auxin as before filtration for 30 minutes. Finally, yeast was vacuum filtered a third time and resuspended in fresh YPD for S phase progression.

For ChIP samples, α-factor cells were crosslinked with formaldehyde and collected as previously described (Wal & Pugh 2012).

Spot Assays

Yeast was grown in YPD or as otherwise specific overnight at 30°C. Log phase cells were collected by brief centrifugation, washed once with PBS, then spotted onto the specified plates starting at $4x10^6$ serially diluted 5-fold.

Western Blots and Quantitations

Standard western blot techniques were used. Tubulin was probed with monoclonal anti-tubulin antibody (Sigma) and GFP was probed with the monoclonal antibody JL8 (Takara). Quantitation of gels was carried out using ImageJ (NIH, Maryland, USA).

ChIP Experiments

ChIP experiments were carried out as described previously with modifications noted below (Wal & Pugh 2012). MNase was titrated for each sample to determine concentrations that would yield ~80-90% mononucleosomal digestion. Immunoprecipitation was carried out using 95% of the MNase digested sample and 4 uL of anti-MCM2-7 polyclonal antibody (UM174, gift from Bell lab, MIT). The remaining 5% of MNase digested sample was processed as 'input' and was treated with Proteinase K and RNase A to prepare libraries for deep sequencing.

Flow Cytometry

0.25 ODs of cells were washed once with water, resuspended in 250uL RNAse A solution (100ug/ml RNAse A, 50mM Tris pH 8.0, 15mM NaCl), and incubated overnight at 37°C. Cells were then centrifuged the pellet was resuspended in Proteinase K solution (125ug/ml Proteinase K, 50mM Tris pH 8.0) for 1 hour at 50°C. The treated samples were centrifuged, resuspended in 1mL staining solution (1uM Sytox green –ThermoFIsher, 50mM Tris, pH8.0), sonicated briefly using a microtip sonicator, and analyzed for flow cytometry using a Guava easyCyte instrument.

Sync-seq replication Timing Experiments

Experiments to monitor replication timing consisted of sample collection to measure population movement through S phase by flow cytometry as well as genome wide replication timing using deep sequencing (Müller et al. 2014). To do so, 3 ODs of cells were collected for each time point and arrested by the addition of sodium azide (0.1% final) and EDTA (20mM final). From these samples, 0.25 ODs were used for flow cytometry analysis and the rest was used for genome wide copy number analysis by deep sequencing.

To prepare genomic DNA for deep sequencing, cells were lysed as for ChIP experiments. Lysates were treated with proteinase K and RNAse A, followed by two consecutive extractions with Phenol-Chloroform-IAA (25:24:1, Fisher). DNA was purified by ethanol precipitation and resuspended in 135uL water in preparation for shearing. A Covaris machine was used to shear the DNA to an average size of ~200bp, following manufacturer's instructions. DNA was then purified by DNA Clean and Concentrator Columns (Zymo Research) before library-making.

Library Preparation and Sequencing

Libraries were prepared using Next Ultra II kits (NEB) following the manufacturer's protocol. Following library preparation, ChIP samples were purified as in (Wal & Pugh 2012) while input samples were purified using a 2:1

ratio of AmpureXP (GE) beads to library. Following PCR amplification, all samples were purified using AMpure XP beads with 0.9:1 beads to PCR ratio. Samples were sequenced using the Illumina NextSeq 500 platform with paired ends.

Data Processing

For ChIP experiments, sequencing reads were mapped to SacCer3 using Bowtie1 using standard parameters and a 650bp upper limit cutoff. V plots and ACS profiles were generated using in house scripts generously shared by Nils Krietenstein (Rando Lab). Coverage files were generated using Deeptools 3.0.2 using 1kb windows around origins of replication, unless otherwise stated.

For galactose induction experiments, coverage files were normalized to the total number of reads mapped to the S. pombe spike in control in 25bp windows. Input signal was then subtracted from the normalized ChIP coverage files to account for background oscillations in the data.

For auxin degradation experiments, coverage files were normalized to counts per million (CPM) and to the non-origin background signal in 25bp windows. Input signal was then subtracted from the normalized ChIP coverage files to account for background oscillations in the data.

For replication timing experiments, coverage files were generated using LocalMapper scripts and Trep was calculated at 1000bp bins genome wide using

Repliscope. Both packages were generously shared by Dzmitry Batrakou (Nieduszynski Lab) and are available at https://github.com/DzmitryGB/. The method is thoroughly described in (Müller et al. 2014). Trep data generated from Repliscope was Loess smoothed in windows of ~50kb using Igor (Wavmetrics, Lake Oswego, OR, USA). At least 6 samples covering a range in S phase progression were used to estimate T rep.

CHAPTER III: Discussion

1. Summary of results

The mechanisms involved in establishing replication timing in eukaryotes remain one of the more elusive aspects of DNA replication. In this project, we explored the effects of modulating levels of the main replicative helicase, MCM, to test models of how multiple MCM are loaded at origins and how differential loading of MCM across budding yeast origins may contribute to the replication timing program.

We found that overexpression of MCM2-7 has no effects on cell viability, MCM abundance, or replication timing. In contrast, reduction of the functional pool of MCM hexamers through the degradation of Mcm4 led to reduced viability and increased sensitivity to replication stress. In addition, reduction of the available pool of functional MCMs during G1 disproportionally affected loading at origins which normally load low levels of the helicase. Lastly, reduction of MCM abundance at origins induced by Mcm4 degradation correlated with delays in replication timing.

Together, these results point to a model where MCM is loaded at origins genome wide based on the activity of origins and their ability to recruit ORC and load MCM. Reducing cellular MCM levels in this model causes reduction of MCM levels at all origins. In addition, our data suggests that reducing MCM pools disproportionally affects low activity origins and further delays their replication timing in S phase. Consequently, these results lend further support to the

multiple MCM model, which posits that differential loading of multiple MCMs at origins contributes to the replication timing program.

2. Experiments to further test the role of MCM in replication timing

The galactose overexpression experiments presented in Chapter II indicate that overexpression of the MCM2-7 hexamer by itself does not lead to changes in MCM loading or replication timing. As discussed in Chapter II, this result could be due to MCMs presence in already high levels in budding yeast cells. Given MCMs tight functional association with Cdt1, however, it would be interesting to test the consequences of overexpressing the heptamer in G1 to gauge any differences in helicase loading and replication timing. Overexpression of Cdt1 would be expected to further increase the available pool of functional MCM hexamer in the nucleus, thus providing an abundance of ready to load MCM hexamers, possibly increasing MCM loading at origins of replication genome wide or at specific loci.

The experiments presented in this thesis used T rep as a measure for the replicating timing of origins. Although T rep is a good approximation of origin activity, it does not take into account passive replication of origins. For example, an origin that loses its MCM signal may experience a concomitant decrease in activity. If this origin is located in an origin poor region of the genome, the change in replication timing would be captured by Trep. However, if this origin is located next to a separate, MCM-rich and early-firing origin, the functional effect of

reduced MCM levels would not be captured by using T rep as a measure of replication timing. An example of this occurrence in our data can be seen in Figure 3.1. After Mcm4 degradation, the levels of MCM by ChIP-siq are reduced. Consequently, the ARS414 peak in the t rep profiles is also lost. However, ARS414 is next to ARS 415 which has a relatively very early t rep and seems impervious to Mcm4 reduction. As a result, the calculated replication time of ARS414 is much earlier than accounted by its activity.

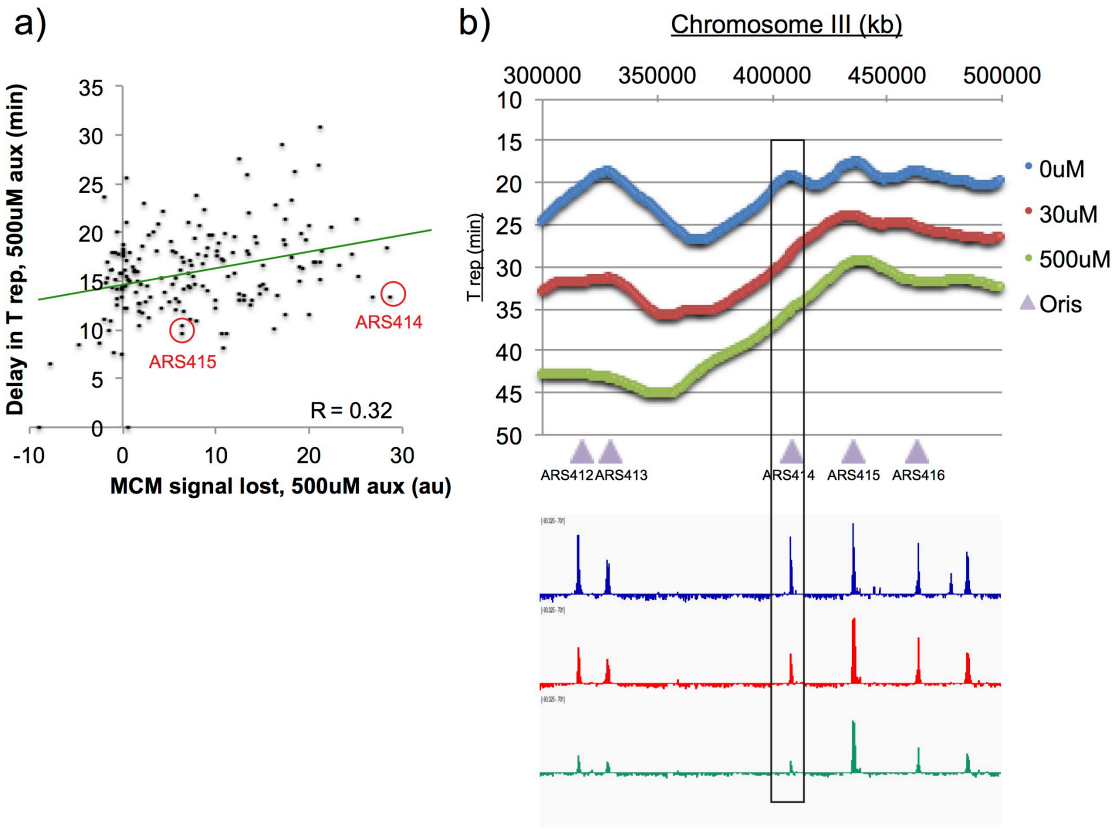


Figure 3.1: Passive replication of origins produces abnormally early t rep values relative to their activity.
a) Plot displaying the correlation between loss of MCM levels and delay in t rep (also shown in Figure 2.15d. The origins discussed in the text are circled in red. b) MCM ChIP-seq and replication timing profiles for untreated and auxin-treated samples, with the apparent outlier from panel a) - ARS414 - highlighted in the box.

A better replication timing parameter is ‘n’, which takes into account an origins time of replication as well as the chances that it may be replicated by a proximal origin (Yang et al. 2010). Although a dataset with n values for a number of origins exists, those values are derived from low-resolution replication timing datasets. In addition, absolute replication timing varies widely based on a number of different factors, including differences in temperature, nutrition, or chemical treatments. Therefore, deriving n for each replication timing dataset may improve the correlation between MCM abundance at origins of replication and replication timing.

One of the questions that emerge from work in this thesis concerns the mechanism by which some origins are resistant to changes in MCM levels and other are sensitive. Analysis of ACS sequences for origins that were sensitive or resistant to changes in MCM levels did not reveal any motifs that were conserved between origins. Recent results indicate that chromatin-remodeling factors have a strong effect on origin architecture, licensing, and activity in vitro (Azmi et al. 2017). It would be interesting to establish a chromatin-binding atlas around origins of replication in G1 arrested cells that encapsulates the full breadth of chromatin factors and modifications at origins of replication. Such studies may shed light on the differences in origin licensing throughout the genome. However, this promises to be a challenging feat that likely involves a delicate balance between many chromatin factors with often opposing roles.

3. Implications of the findings in disease

As discussed in Chapter I, regulation of MCM levels and replication timing has important implications for disease. In many cancers, overexpression of MCM proteins has become a leading diagnostic and prognostic factor (Giaginis et al. 2010). Experiments from this thesis showed that overexpression of MCM2-6 did not have any effects in budding yeast. However, lowering the levels of MCM helicase had drastic effects on where MCM was loaded and the replication timing profile, showing that altered MCM pools can have strong implications for cellular function. Cancerous states have been shown to have large changes in their replication timing program (Blumenfeld et al. 2017). Testing whether MCM loading is involved in the establishment of these altered timing programs may yield new insights into how these cancers arise and how they propagate.

Meier-Gorlin syndrome is a dwarfism disorder characterized by mutations in various pre-RC components, including ORC1, ORC4, ORC6, CDC6, and CDT1 (Bicknell et al. 2011a). Disease-associated mutations in ORC1, for example, have been shown to reduce chromatin binding of the complex, reduce pre-RC assembly, and impair activation of replication origins in a similar fashion as reduction of MCM5 levels (Bicknell et al. 2011b). Results from experiments in this thesis indicate that the ability to recruit ORC and load MCM is also crucial for the maintenance of proper levels of origin licensing in budding yeast. Importantly, they suggest that one of the key consequences of reduced MCM loading

concerns its effect on replication timing. Whether a similar mechanism modulates MCM loading and affects replication timing in the Meier-Gorlin disease model is unclear. Therefore, understanding the mechanistic consequences of having low MCM in budding yeast could shed light on how ORC1 mutations contribute to disease in Meier-Gorlin syndrome.

4. Looking to the future

Several advances in technology and biochemical techniques have made possible the testing of biological systems at an unprecedented level of detail. The use of Chromatin Endogenous Cleavage and sequencing (ChEC-seq), for example, promises to shed a new level of detail on the DNA footprint of DNA binding factors such as MCM (Zentner et al. 2015). This technique uses the fusion of a protein of interest to MNase to digest DNA specifically around the binding site of a protein and very high signal-to-noise ratios. Recent experiments have used ChEC-seq to monitor MCM binding genome wide and data from specific origins indicates that only one well-positioned MCM hexamer is present in G1 (Foss et al, BioRxiv 2019). How these trends will hold up at all origins remains to be seen. Using this technique, several additional questions may be addressed at a higher level of resolution, including the binding dynamics of ORC, MCM, and nucleosomes around origins of replication.

The recently developed system of in vitro fully reconstituted replication in budding yeast is another exciting frontier that will yield unprecedented

mechanistic detail into a number of questions (Yeeles et al. 2015). Using this system, it was determined that the histone chaperone FACT was required for replication of chromatinized DNA templates. Using this system, additional chromatin remodeling factors can be tested to gauge their specific effect on origin licensing and activation, which has important implications for MCM loading.

As described throughout this thesis, one of the aspects of replication timing that is still being debated is whether more multiple MCM double hexamers are loaded in budding yeast origins in vivo. In vitro experiments have shown that loading of multiple MCMs is at the very least possible (Remus et al. 2009; Ticau et al. 2015). High-resolution genome-wide studies, on the other hand, indicate that at most one MCM double hexamer is loaded onto origins of replication (Belsky et al. 2015). In order to precisely answer this question, and to grasp the distribution and variety of MCM loading in budding yeast, new assays must be developed to measure the number of helicases loaded at single origins in an in vivo context. Single-molecule TIRF microscopy promises to be an exciting avenue to pursue these questions. As part of my thesis research, I proposed and made several advances in trying to measure the number of MCMs loaded at single replication origins by TIRF microscopy (see Appendix C for results and discussion).

5. Concluding remarks

The work presented in this thesis furthers the current understanding of how replication timing is established in budding yeast and the role that the MCM helicase plays in that process. Data presented here suggests that, at origins spanning the budding yeast genome, changes to MCM loading correspond to changes in replication timing. Importantly, this work also sheds light on the process of MCM loading during G1, showing that origins of replication are disproportionally affected by changes to MCM levels and that these changes have strong implications for how S phase proceeds.

APPENDIX A: Studies of Real-Time, In-Vitro Replication in Frog Egg Extracts Using PhADE

A.1. Introduction

The question of how cells deal with DNA damage is a major focus of research that directly links the basic science understanding of DNA replication and its clinical implications in disease. DNA damage can originate from a variety of sources. Those external to our bodies include the ultraviolet (UV) rays of the sun, environmental toxins, or radiation from common medical technology. In addition, our own bodies produce reactive byproducts as a result of normal cellular metabolism. These substances can covalently modify DNA to make mutagenic lesions. Lastly, the mere process of DNA replication, though impressively accurate, has an inherent error rate. The error rate for replicative DNA polymerases is on the order of 1 in 105 and is often exacerbated by hard-to-replicate genetic regions containing secondary DNA structures, repetitive sequences, or bound proteins (Thomas et al. 1991; Loewen & Switala 1995).

If damaged DNA is sensed in G1, G2, or M phases of the cell cycle, the cycle is arrested by DNA damage checkpoints until the damage is repaired (Sancar et al. 2004). However, the circumstances are more complicated during S phase due to ongoing DNA replication. As a result, the S-phase DNA damage checkpoint slows down replication of the genome and activates the appropriate pathways to deal with the damaged DNA (Labib & De Piccoli 2011). These pathways include the up-regulation of dNTP pools by stimulation of ribonucleotide reductase (RNR) activity and expression, transcription of replication and repair factors involved in

the damage response, inhibition of origin firing, and stabilization of replication forks (Labib & De Piccoli 2011). Of these four responses, the inhibition of origin firing and stabilization of replication forks are responsible for the overall slowing of replication that is observed in yeast and mammalian cells in response to DNA damage (Painter & Young 1980; Paulovich & Hartwell 1995; Tercero et al. 2003).

Important questions remain in the field about how replication forks slow in response to damage, and whether this behavior is mediated by the activation of the S-phase damage checkpoint. Studies in mammalian cells have shown that replication forks slow in response to damage in a checkpoint dependent manner (Merrick et al. 2004). However, studies in yeast have suggested that replication fork slowing in response to DNA damage is independent of the checkpoint, suggesting a model where slowing mediated by replication forks is solely due to replication forks physically interacting with damage (Tercero & Diffley 2001; De Piccoli et al. 2012; Iyer & Rhind 2013). The data supporting a crucial role for the checkpoint in mediating replication fork dynamics in response to DNA damage is considerable (Merrick et al. 2004; Seiler et al. 2007; Kumar & Huberman 2009). Nonetheless, the mechanisms of checkpoint-mediated fork slowing are still mysterious.

From in vitro biochemical experiments, we know that lesions such as those caused by treatment with the DNA methylating agent methyl-methane sulfonate (MMS) block DNA polymerase progression (Larson et al. 1985; Wyatt & Pittman

2006). However, what happens to the replication fork once one of its polymerases stall is not well understood. If the damage is encountered on the lagging strand, it is conceivable that a gap would be left between the site of the stall and the upstream Okazaki fragment and the fork would continue on, unaffected. Damage encountered on the leading strand is more complicated because the replicative helicase can continue to unwind DNA after the polymerase has stalled, exposing tracts of single-stranded DNA. The single-stranded DNA is bound by RPA and initiates checkpoint activation (Byun et al. 2005; Friedel et al. 2009). The active checkpoint then phosphorylates numerous proteins associated with the fork and averts additional separation of the helicase and the stalled polymerase (Ohouo et al. 2010; Lossaint et al. 2013). Thus, it remains unclear how the checkpoint facilitates restart of the stalled polymerase. Translesion polymerases, strand-switching, replication-coupled recombination, and down-stream repriming of the leading-strand polymerase are mechanisms that may be involved (Ciccia & Elledge 2010). However, how the checkpoint may regulate some or all of these mechanisms is unknown. Moreover, the fact that in the absence of the checkpoint forks seem to be able to move through sites of DNA damage without slowing down suggests that the checkpoint-dependent response to damage is slower than the default response (Kumar & Huberman 2009). In addition, it is unknown whether checkpoint activation leads to global slowing of all forks, regardless of whether they've encountered damage, or if the

checkpoint-dependent slowing only affects replication forks that encounter damage on their templates (Iyer & Rhind 2013).

The biochemical complexity of DNA replication has made mechanistic in vitro studies difficult, and the limits of studying dynamic structures using population experiments have obscured how individual replication forks respond. In order to overcome these challenges, one can analyze the response of individual replication forks to DNA damage in real-time using a technique called PhotoActivation, Diffusion, and Excitation (PhADE) in the biochemically tractable system of Xenopus laevis egg extract system (Loveland et al. 2012). This assay makes it possible to study individual replication fork dynamics in real-time in a vertebrate system. Using PhADE, one should be able to determine replication fork rates on various damaged λ phage DNA structures under checkpoint competent and inhibited conditions. Several features of this approach allow for the study of replication forks at an exceptional level of detail.

First, the Xenopus egg extract system is biochemically tractable and able to recapitulate regulated DNA replication in a vertebrate system (Garner & Costanzo 2009). Techniques for the biochemical manipulation of these extracts have been refined to allow for immunodepletion of specific proteins and addition of exogenously expressed factors (Lebofsky et al. 2009). Moreover, preparation of extracts that reproduce G1 and S-phase regulation of origin establishment and activation, allows for reconstitution of fully active replication forks (Walter et al.

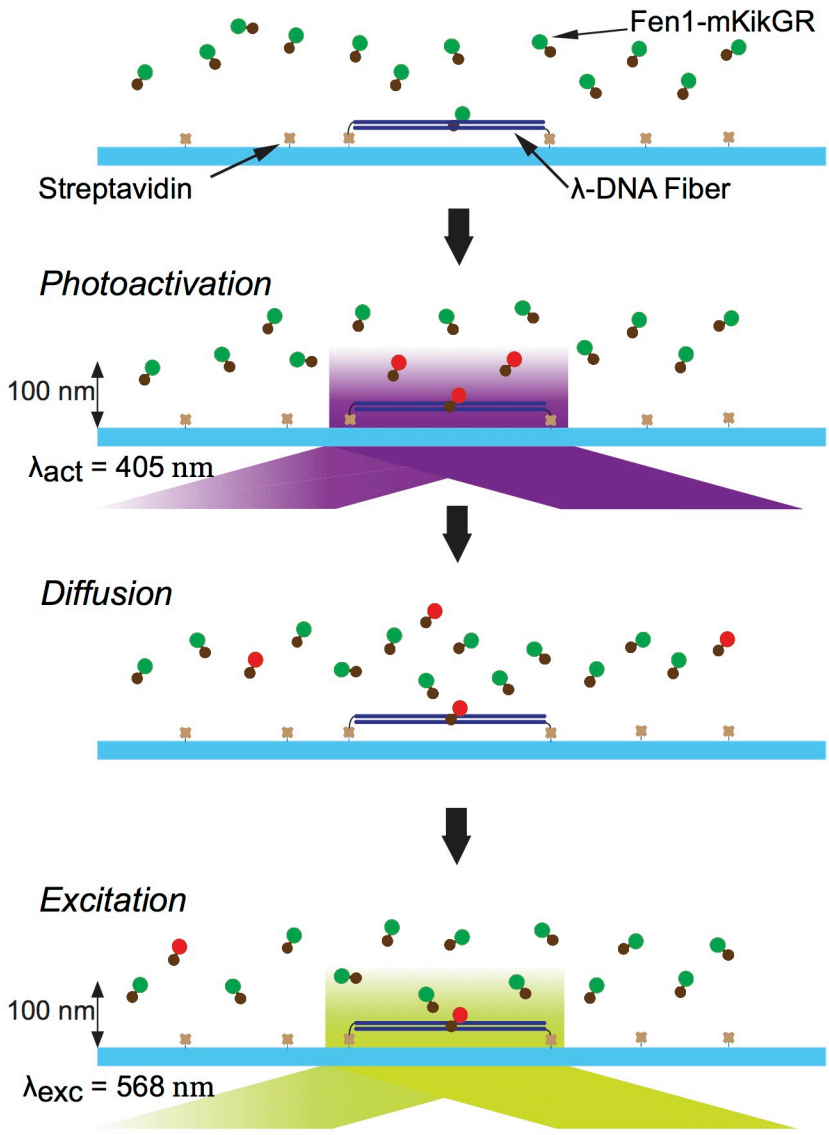
1998). In particular, high speed supernatant (HSS) is a G1 cytoplasmic extract of Xenopus eggs that, when incubated with exogenous DNA, such as λ phage chromosomes, allows for the establishment of pre-RCs. Nucleoplasmic extract (NPE) is an S-phase extract that provides the high CDK/DDK concentration that is required for the start of DNA replication, and also effectively limits replication to only one round by preventing re-association of MCM with chromatin.

Second, λ phage DNA is an excellent template with a known sequence that is highly customizable. The genome of λ phage is 48.5 kb in length, is readily isolatable in large quantities from bacteriophage lambda, and contains unique 12-bp single-stranded 5’ overhangs at each end. These ends can be annealed to complementary single-stranded oligos in different combinations and functionalized with 3’ biotin molecules to provide templates for microfluidic flow cell experiments and single-molecule analysis (Yardimci et al. 2012).

Third, PhADE allows for the visualization of single-molecules at physiological concentrations of proteins. Regular single-molecule fluorescence microscopy relies on a low concentration of fluorescent molecules in order to reduce background signal. Total Internal Reflection Fluorescence (TIRF) microscopy can reduce background fluorescence by selectively exciting molecules that are within 100-200 nm of the surface. However, the maximum concentration of fluorescent molecules in TIRF experiments cannot exceed ~10 nM without significantly increasing background signal. This constraint poses a problem for the study of

replication proteins, which are present in Xenopus egg extracts in the micromolar range. PhADE circumvents this problem by fusing a protein of interest to a photoconvertible fluorophore.

In PhADE, a protein of interest is fused to the photoconvertible protein mKikGR and introduced into a flow cell containing a functionalized surface. The fusion protein binds to the surface through interaction with a substrate or binding partner and becomes immobilized. Unmodified mKikGR fluoresces green when excited with 488 nm light (mKikG). Excitation with 405 light irreversibly converts the protein to a red fluorescent form that is excitable by 568 nm light (mKikR) (Habuchi et al. 2008). TIRF is used to photoconvert only mKikG molecules that are close or bound to the surface. Diffusion reduces the concentration of free, photoconverted mKikR in the TIRF layer as these are exchanged for non-photoconverted mKikG fusion proteins. Excitation at 568 nm then allows for visualization of single molecules of mKikR that are bound to the surface of the flow cell (Figure A1). I used a fusion of the flap endonuclease Fen1 to mKikGR (Fen1-mKikGR) to visualize replication fork movement in real time. Fen1 is involved in Okazaki fragment processing during replication in Xenopus egg extracts and has been shown to travel with the replication fork, presenting a suitable way of tracking replication forks in real-time (Burgers 2009; Loveland et al. 2012).



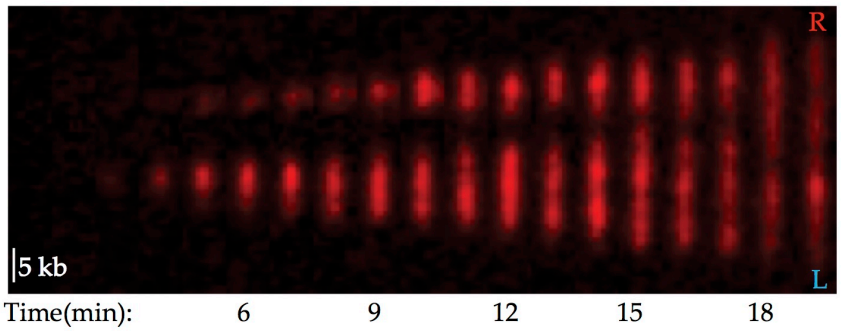
Adapted from Loveland … van Oijen (2012) Nat Methods

Figure A.1 - PhotoActivation, Difussion, and Excitation (PhADE) experiment scheme. The photoactivatable fusion protein Fen1-mKikGR associates with active replication forks on λ phage DNA molecules that have been doubly-tethered to the coverslip. The λ phage DNA-bound Fen1-mKiGR is photoactivated using 405nm light within the TIRF volume of the flow cell, while unbound molecules diffuse away. The photoactivated protein is then imaged 568nm light to monitor replication fork location on the λ phage DNA molecule.

A.2. Results

Using PhADE, the goal of this project was two-fold. First, I wanted to measure the dynamics of single replication fork progression on undamaged and damaged DNA templates. Single-molecule work done on bacterial helicases shows that DNA unwinding proceeds at a variety of speeds that are independent of each other and of previous speeds measured for the same helicase (Liu et al. 2013). Similar data has been difficult to acquire in higher eukaryotes due to the complexity of the biochemical system. In PhADE, this analysis can be accomplished by tracking uniaxial progression of photoactivated Fen1-mKiKGR (i.e. the replication fork) as a function of time (Figure A.2). Whereas in vivo Fen1 travels with the fork, the higher concentration used in PhADE causes Fen1 to also be deposited behind the fork. Replication forks traveling on undamaged DNA were expected to move most of the time without pausing. Indeed, PhADE experiments showed that replication forks proceed on undamaged templates at an average speed of 484 bp/min (Figure A.3a). This data correlates well with in vivo rates measured from Xenopus embryos, as well as estimated replication rates in extracts (Mahbubani et al. 1992; Hyrien & Mechali 1993). In addition, comparison of fork speeds emanating from the same origin indicated that the replication speeds of sister forks are independent of each other, consistent with previously published data (Figure A.3c) (Yardimci et al. 2010).

a.



b.

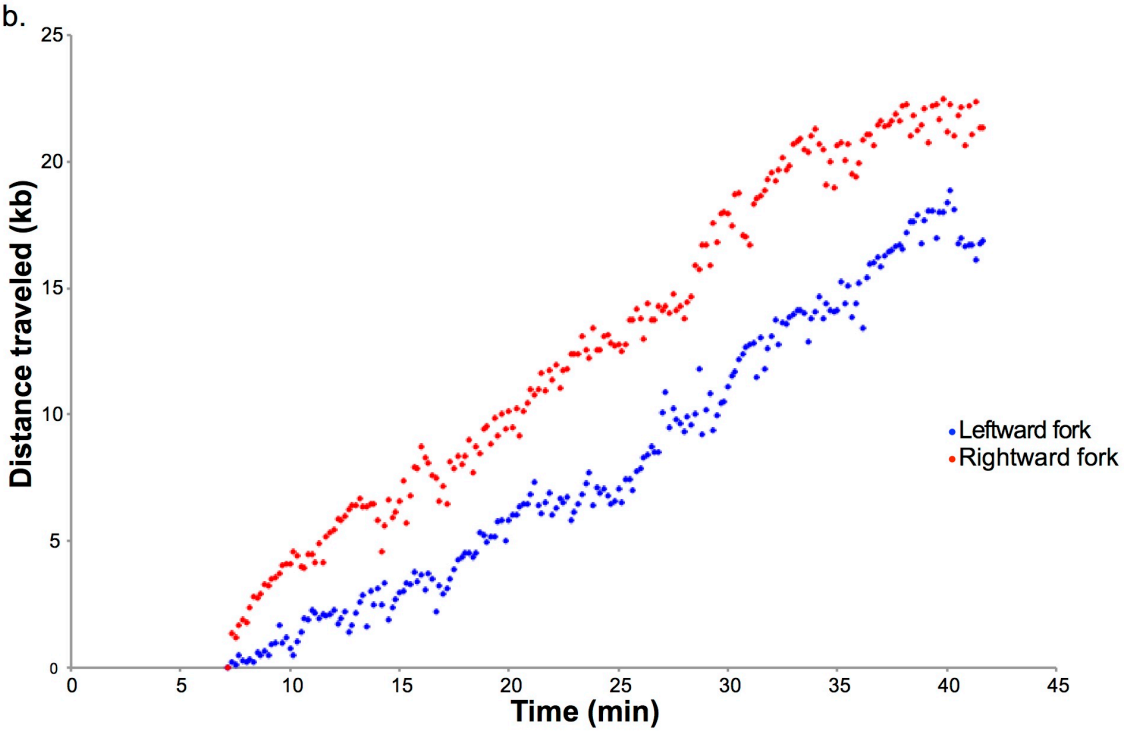


Figure A.2: Monitoring replication fork movement using PhADE
Representative kymograph (a) and plot (b) of replication fork analyses.

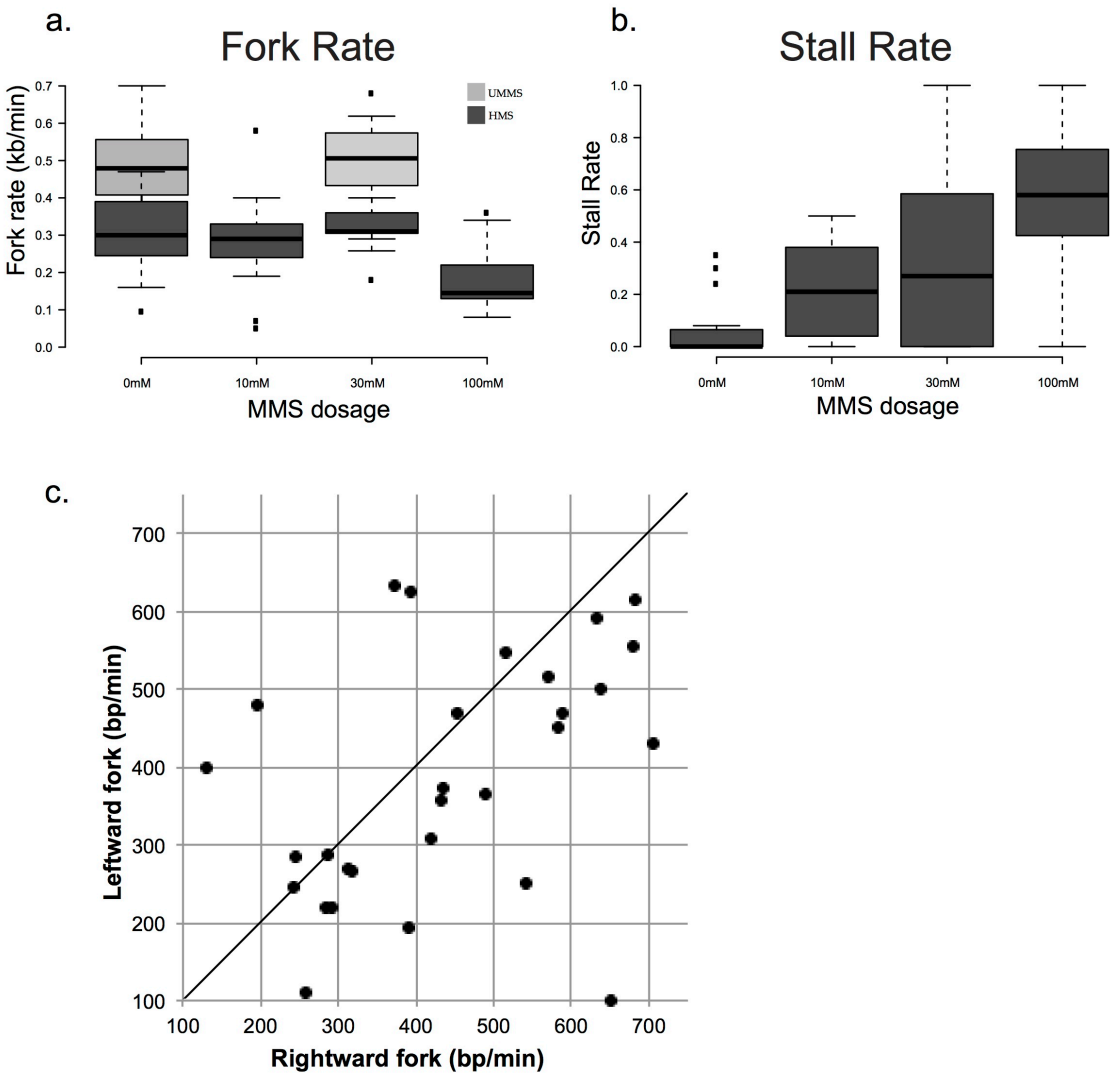


Figure A.3 - PhADE data.
a) Fork rates from -undamaged and MMS-damaged λ phage DNA molecule experiments. b) Stall rates of replciation forks from λ phage DNA molecules treated with varying concentrations of MMS (data obtained by Nick Rhind at Harvard Medical School). c) Replication fork speeds of sister forks originating from the same replication origin.

It has been well established that bulk replication proceeds slower in the presence of various forms of DNA damage (Tercero & Diffley 2001; Merrick et al. 2004; De Piccoli et al. 2012; Iyer & Rhind 2013). However, the single replication fork dynamics that add up to produce these population behaviors have largely remained untested due to the technical challenges of performing such

experiments. One can imagine a scenario where the speed of replication forks, while moving, is similar on damaged and undamaged templates (Figure A.4a vs. A.4b). In this scenario, the observed replication slowing could be a result of fork stalls, as replication forks encounter and deal with DNA. Conversely, it may be possible that the presence of DNA damage elicits a global response, possibly in the form of a phosphorylation cascade mediated by the checkpoint, which acts on all replication forks to slow them down (Figure A.4c). Using PhADE, I wanted to investigate replication fork dynamics in the presence of MMS-mediated DNA damage. Preliminary data using λ phage DNA templates that were treated with 30mM MMS suggested that replication forks maintained a speed of 481 bp/min while moving (Figure A.3a). This number is very similar to that observed for undamaged templates, suggesting that stall rates are the main contributing factor to DNA replication slowing in the presence of damage in frog egg extracts. Indeed, unpublished data obtained by Nick Rhind suggests that, for lower doses of MMS, replication fork rates remain constant while stall rates (defined as fraction of time spent stalled) increase with increasing concentration of DNA damage (Figure A.3b).

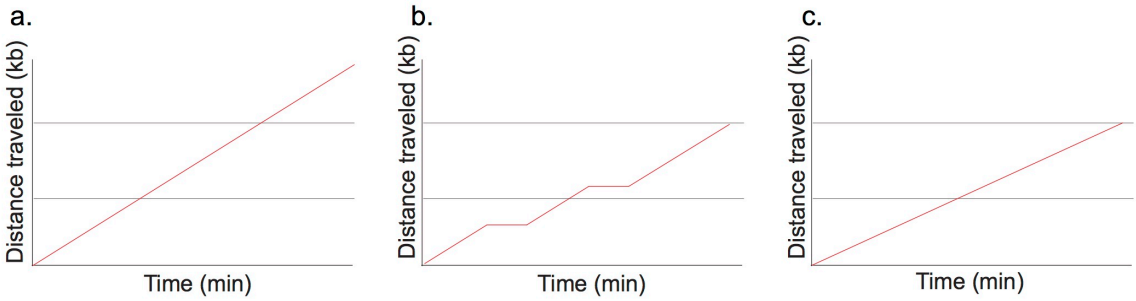


Figure A.4: Possible scenarios for replication fork movenent on MMS-damaged and -undamaged DNA.
a) Replication forks on undamaged DNA travel at constant speeds. Replication forks on damaged DNA may travel at similar speeds but stall more often (b) or they may travel at slower speeds (c).

The second goal of this project was to explore the role of the S-phase checkpoint in mediating replication fork progression in the presence of damage. Activation of the S-phase checkpoint in response to DNA damage is important for the slowing of overall replication(Labib & De Piccoli 2011). Phosphoproteomic analyses have shown that many proteins directly associated with active replication forks are phosphorylated in response to DNA damage. This phosphorylation is checkpoint dependent, relying on the activity of the main checkpoint kinase ATR (Lossaint et al. 2013). However, the details of how checkpoint-dependent phosphorylations affect replication fork dynamics are unclear. Although slowing of DNA replication in budding and fission yeast seems to be independent of the checkpoint, data from mammalian cells and other organisms support a crucial role for the checkpoint in mediating replication fork dynamics in response to DNA damage, including replication fork slowing (Merrick et al. 2004; Seiler et al. 2007; Kumar & Huberman 2009).

To probe the role of the checkpoint in mediating replication fork dynamics, I planned to perform PhADE experiments on MMS-damaged and undamaged templates using checkpoint-capable or -inhibited extracts. The small-molecule inhibitor VE-821 has been shown to be an efficient inhibitor of the main checkpoint kinase ATR and would efficiently inhibit the checkpoint response (Reaper et al. 2011). However, technical difficulties with various aspects of performing PhADE prevented further data collection.

Several factors contributed to the lack of further progress on this project. The major hindrance rested on the variability of frog egg extract preparations, specifically NPE. Some preparations produced extracts that allowed for fast replication, as tested by bulk 32P-dATP incorporation during plasmid replication, whereas others were considerably slower. The factors that contribute to the quality of the extracts are believed to include the quality and quantity of the frog eggs, which are dependent on the health and use of frogs. Furthermore, the quality of the purified frog sperm chromatin that's necessary for nuclei formation and NPE extraction is a determinant of the yield and quality of NPE (Lebofsky et al. 2009). Lastly, the speed and technique of preparation of each batch of extracts during the necessary 6-8hr protocols often leads to significant difference in extract quality between different preparations from the same person and also from different people.

In PhADE, these differences seemed to be exacerbated, as extracts that weren't as efficient at bulk replication often completely failed at producing measurable data by PhADE. In addition, the difference in extract quality likely explains the ~150 bp/min difference in replication fork speeds of undamaged templates between experiments performed at Harvard Medical School (HMS) by Nick Rhind and those performed at UMass Medical School (UMMS) by myself (Figure A.3a). It is important to note that, while the kinetics of replication may be slightly different, the underlying biology between the different experiments is largely the same (personal communications from Nick Rhind and Walter lab, HMS).

An additional hindrance in continuing data collection using PhADE at UMMS was the preparation of high quality microscope slides. Slide preparation is an infamously stubborn step in performing successful single-molecule experiments (Visnapuu et al. 2008). In experiments that use purified proteins, the problem often manifests itself as increased background, which makes it harder to discern single molecules. In PhADE, sub-optimal slides led to a "stickier" background, which caused replication of λ phage DNA to stop prematurely even without DNA damage and for fibers to non-specifically bind to slide surfaces. By trying various methods of slide preparation, this problem was largely eliminated. However, there remained a large variability in the quality of replication using the same batch of slides with different extracts. Put together, these challenges did not make it feasible to move forward with the PhADE project.

A.3. Discussion and Future Directions

Replication forks are complicated machines that are central to the successful duplication of DNA and the maintenance of genomic stability. Ensemble data gives an averaged view of how fork proteins are modified in response to damage and how overall replication kinetics are affected. However, ensemble data cannot address the heterogeneity of replisome function and dynamics. Single-molecule experiments promise to reveal the complexity of fork interactions with damage—for instance, whether and how often they slow, stall, or arrest—and provide tools to investigate the biochemical mechanisms by which these interactions are regulated.

Using PhADE, I investigated the replication dynamics of single replication forks on undamaged and MMS-damaged DNA templates. Experiments using undamaged templates revealed replication fork speeds similar to those estimated from static methods, confirming the assay as a valid means of measuring replication fork dynamics. Furthermore, they showed that replication forks emanating from the same origin behave independently of each other. Experiments using MMS-damaged DNA templates showed that average replication fork speeds for moving forks do not change even when the DNA is damaged (Figure A.3a). Although the quantity of the data that I collected precluded the estimation of stall rates, data from Nick Rhind suggests that stall rates increase with increasing MMS-damage (Figure A.3b). Together, these data

point to a model where replication slowing in response to MMS-mediated DNA damage in frog egg extracts arises from multiple stalling events, as opposed to a global modification of replication fork speeds.

It is important to note that the spatial resolution of the assay is about 1 kb and the temporal resolution is 10 seconds. These limits mean that it's not possible to see individual stalling events on templates with more than 1 lesion/kb or stalling events that last less than 10 seconds. Instead, such sub-resolution events are expected to be observed as uniform slowing of replication forks. Since the data only shows uniform slowing occurring at 100mM MMS, it appears that sub-100mM MMS causes damage that's within the resolution of the assay. In addition, since the average fork rates between treated and untreated DNA do not change up to 30mM MMS, the 10s timescale of the assay seems to be able to resolve stalling events, which increase as MMS concentration is increased. Whether and how the replication checkpoint is involved in establishing these dynamics remains to be tested.

Numerous factors contributed to unfeasibility of continuing the use of PhADE at UMMS. The variability in the quality and quantity of frog egg extracts that I prepared at HMS and the logistical hurdle of the collaboration made troubleshooting much more challenging. Combined with the variability and relative unpredictability of microscope slide functionalization, continuing this project was not feasible. In order to circumvent the problems rising from

variability in frog egg extracts, it may be possible to perform large numbers of NPE preparations, discard those that are not satisfactorily efficient at replication, and combine the ones that are efficient. This method would yield large amounts of equally efficient NPE, which would allow for direct comparison between different experimental setups on different days and weeks. However, due to the fact that our lab does not have it's own frog facility, it was not logistically possible to use this approach.

Despite technical hurdles in setting up a successful PhADE assay, doing so promises to yield interesting insights into the nature of replication fork dynamics. The following are future directions for a robust PhADE assay. First, it would be interesting to probe the replication fork dynamics under different forms of damage, such as ultraviolet radiation (UV). UV primarily causes pyrimidine dimers, 6-4 photoproducts, and oxidation at C8 of guanine bases(Cadet et al. 2012). These lesions stall polymerases, activate the ATR-mediated DNA damage checkpoint, and cause slowing of overall replication(Guo et al. 2000). This treatment represents a directly relevant source of DNA damage in humans, as UV radiation-induced damage caused by exposure to the sun's rays is one of the most direct causes of cancer. How checkpoint activity affects individual replication forks that encounter UV damaged DNA remains unclear.

DNA damaging agents cause randomly-located lesions that inhibit polymerase and fork progression. Although titration of the damage dose is useful

in gauging different replication fork dynamics, it does not provide information about the specific geometry of lesions and how forks encounter them. Two important questions exist. First, does the replication fork behave differently depending on the polarity (leading v. lagging strand) of the lesion? Second, what is the effect that the number and spacing of lesions have on fork progression? To address these questions, it would be interesting to construct λ DNA molecules with strand- and location-specific damage. Protocols for the generation of these molecules have been established and allow for specific probing of how replication forks progress through damaged DNA (Yardimci et al. 2012).

Lastly, another question that remains unanswered in the vertebrate system of frog egg extracts is whether the checkpoint-dependent slowing of replication in S phase is a local or global response to DNA damage(Iyer & Rhind 2013). Checkpoint-dependent slowing can be local if it is due solely to slowing of forks that encounter damaged DNA. Conversely, slowing can be a global phenomenon whereby the encounter of a few replication forks with damaged DNA bases elicits a response that slows all replication forks, regardless of whether they have encountered damage. To address this question, it would be interesting to perform PhADE experiments using a mix of damaged and undamaged λ phage DNA in the presence of an active DNA damage checkpoint. Treated templates in this experiment can be fluorescently-labeled using oligonucleotides, while keeping the untreated templates unlabeled. In doing so, it would be possible to determine

whether checkpoint activation affects replication fork rates on undamaged or damaged templates in the same experiment.

In summary, single-molecule experiments promise to elucidate the finer details of replication fork progression, especially as it pertains to how forks deal with damaged DNA. Although there are numerous technical hurdles, the results would shed light on some of the confounding results seen in population experiments. The vertebrate system of frog egg extracts is an excellent system to perform these experiments because of its closeness to human cells. However, recent advances in fully reconstituted in-vitro replication using yeast proteins will likely also be key to teasing apart the intricacies of eukaryotic DNA replication.

A.4. Materials and methods

Xenopus egg extracts and replication

Xenopus egg extract preparation and replication efficiency tests were conducted as previously published, but using pBS-SKII(+) plasmid as a template (Lebofsky et al. 2009).

Protein Purification

Protein purification of Fen1-mKikGR(WT), Fen1-mKikGR(D179A), and p27-kip were conducted using standard techniques and as previously described (Loveland et al. 2012).

Preparation of λ phage DNA constructs for PhADE

The protocol for producing biotinylated λ phage DNA using Klenow Exo- was provided by Gheorghe Chistol (HMS). Briefly, in a 100ul total reaction, 10ug λ phage DNA (NEB: N3013) was biotinylated using 100uM dATP, dGTP, dTTP each (NEB), 60uM dCTP-biotin (Life Technologies), and 5uL Klenow Exo- (NEB) in 1x Buffer #2 (NEB) at 37°C for 30 minutes. Klenow Exo- was inactivated at 75°C and the reaction mixture was drop dialyzed 3 times against 50 mL 15 mM Tris pH 7.8, 5mM EDTA using 0.025 μm disk filters (Millipore VSWP02500). Biotinylated λ phage DNA was diluted 1:200-400 in ELB++ before being flown into chambers for single molecule experiments.

To prepare damaged λ phage DNA for single molecule experiments, biotinylated λ phage DNA was incubated with 30mM MMS in a total volume of 250uL ELB+ for 30 minutes before being loaded onto flow cells for tethering, followed by immediate washing with ELB++.

Coverslip preparation

Coverslips for single-molecule microscopy were prepared as previously described, with changes noted below (Yardimci et al. 2012). The last step of coverslip functionalization involves rinsing away excess biotin-mPEG from slides with milliQ water. The milliQ water at multiple UMMS stations was measured to be pH 6.5-6.7 on multiple occasions. By comparison, milliQ water at HMS, where

the method was established, consistently measured between pH 8.3 and 8.7. Personal correspondence from other labs using single-molecule microscopy at UMMS pointed to acidic water at this step being a possible culprit for producing non-specific sticking to coverslips (Andrew Franck, Melissa Moore lab). To alleviate this problem, 100mM sodium bicarbonate was used to rinse coverslips after functionalization.

Microscope

Flow cells were loaded on a custom made slide holder and mounted on an inverted microscope (Olympus Ix81) equipped with an Olympus 100x 1.49 NA oil immersion objective, three Hamamatsu EM-CCD digital cameras, and Prior Nanoscan Z and H31XYZE stage controllers. Laser excitation was generated by CellTIRF lasers at 405nm, 491nm, 568nm, and 650nm.

Single-molecule replication reactions

Experiments using λ phage DNA and Fen1-mKikGR(WT) purification (used at 2.65uM final) were conducted as previously published (Loveland et al. 2012). DNA was flown into flow cells at 100ul/min for an average extension of 77% of the contour length of λ phage DNA.

Tracking replication fork movement

Replication bubbles emanating from doubly-tethered λ phage DNA molecules were tracked using ImageJ. The X,Y position of each fork was marked on consecutive frames. This position was corrected for frame shifting by using quantum dots as fiducial markers (Loveland et al. 2012). The number of base pairs per pixel was calculated by dividing 48,502bp (length of λ phage DNA) by the total length of doubly tethered λ phage DNA (in pixels) as determined by Sytox staining (Invitrogen). To obtain replication fork rates, Igor Pro (Wavemetrics, Lake Oswego, OR, USA) was used to obtain a linear fit of the data obtained by ImageJ, with the slope of the fit providing the replication fork speed. Constantly replicating regions were determined on a fork-by-fork basis.

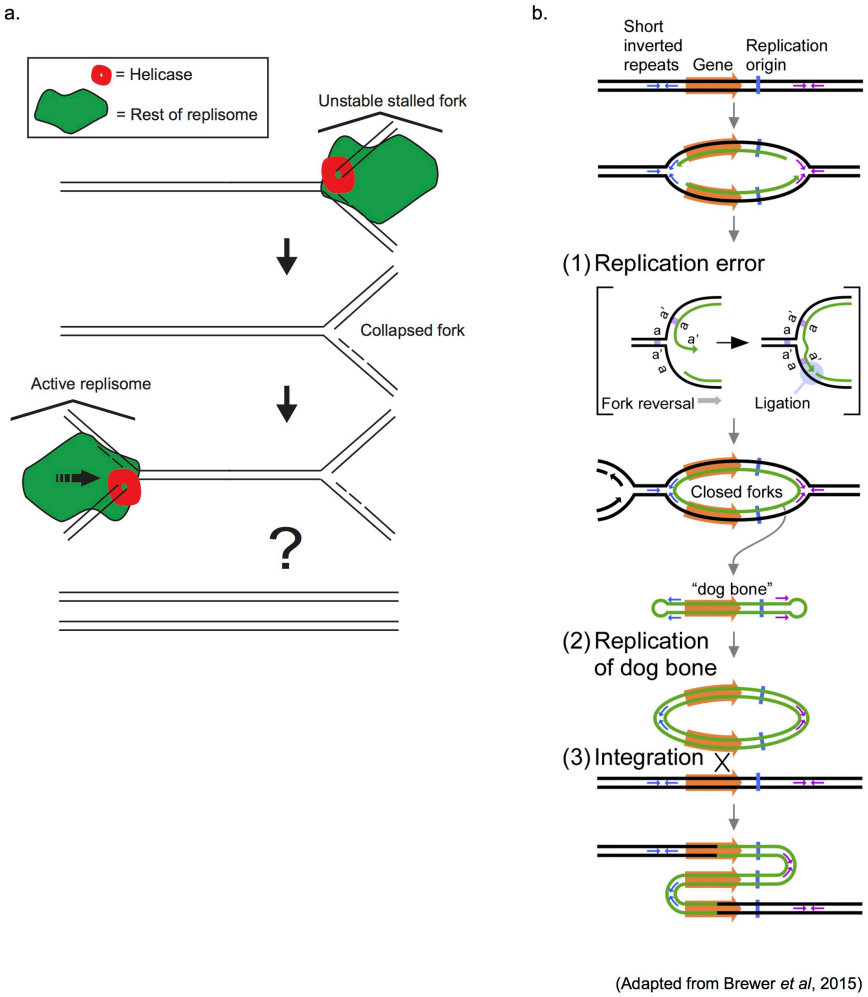
APPENDIX B: Replication Dynamics of Unusual DNA Structures in Frog Egg Extracts

B.1. Introduction

DNA replication forks are susceptible to stalling due to a large number of factors, both endogenous and exogenous (introduced in Appendix A). Although mechanisms exist to maintain the stability of the replisome, various genomic and chemical conditions may lead to breakdown of active forks. Nonetheless, flanking DNA must be replicated in order to avoid gross genomic aberrations. Although adjacent replication forks should easily be able to accomplish this task, their interaction with the DNA structures resulting from a non-functioning fork has yet to be thoroughly explored.

Replication forks can experience a variety of polymerase-stalling events. Depending on the kind and amount of damage, the replication checkpoint may play a crucial role in maintaining a functioning fork, as has been shown in budding yeast (Lopes et al. 2001; Sogo et al. 2002). Lack of a functioning checkpoint can lead to fork reversal and aberrant fork structures that lose their ability to replicate, often referred to as collapsed forks (reviewed in Cortez, 2015). In the event that a collapsed fork is formed, a fork coming from the opposite direction would be able to replicate any unreplicated DNA. If that is not an option due to two adjacent collapsed replication forks, it may be possible for dormant origins that have been shown to fire under replication stress to salvage any unreplicated DNA (Woodward et al. 2006). In either case, an active replication fork would encounter the DNA structure of a collapsed fork (Figure B.1a). How

this encounter is resolved to yield properly duplicated DNA has not been explored to great detail. What structures are produced as a result of this encounter? Are specific proteins involved in ensuring successful resolution of the collapsed fork? In addition, aside from salvaging unreplicated tracts of DNA due to a collapsed fork, this encounter would comprise a replication termination event (introduced in Chapter 1), the fine molecular details of which remain elusive.



(Adapted from Brewer *et al*, 2015)

Figure B.1: a) Active replication forks may encounter collapsed fork structures after replication stress.
How is that encounter successfully resolved? Which factors are involved? b) Model for the generation of inverted repeats through the Origin Dependent Inverted Repeat Amplification (ODIRA) mechanism

Another unusual structure proposed to be encountered during replication is a spanning fork, in which the leading and lagging strands of a collapsed fork have been ligated (Brewer et al. 2011). Spanning forks have been proposed to be intermediates in origin-dependent inverted-repeat amplification, whose specific geometry doesn't fit most models for the generation of repeated DNA sequences. In this model, an active replication fork encounters a ligated fork and prompts the extrusion of the ligated DNA to form a "dog bone" structure, whose ARS-mediated duplication and subsequent integration produces the inverted repeat geometry (Figure B.1b). Transformation of budding yeast with inverted-repeat generating intermediates of this model has shown that their integration is feasible and produces the expected DNA repeat geometries (Brewer et al. 2015). However, it remains unclear how replication forks are able to move past the spanning fork structures successfully and to generate the "dog bone".

In order to address how replication forks deal with the scenarios described above, we devised two DNA structures: one mimicking a collapsed fork ('forked structure') and one mimicking a covalently closed fork ('spanning structure') (Figure B.2a). We then used the frog egg extract system (introduced in Appendix A) to probe the stability of these DNA structures in frog egg extracts as well as their ability to replicate.

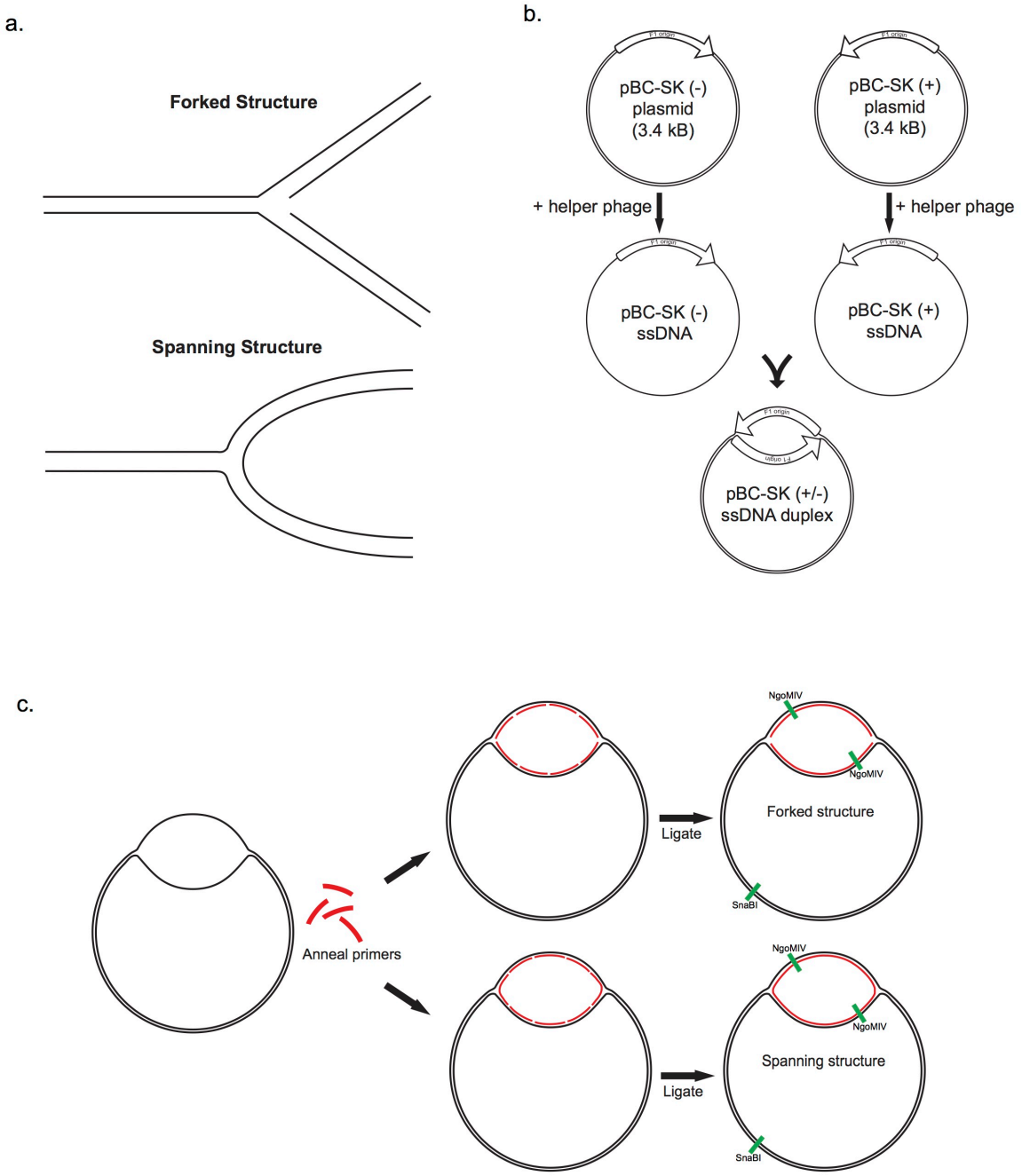


Figure B.2: Forked and spanning DNA structures.
a) Forked DNa structure and spanning DNA strcutres. b) pBC SK( ) and (+) ssDNAs are complementary at all regions except f' orgin. c) Both structures can be completed by annealing primers complementary to f' origin. Primers labeled with 32P-ATP by in vitro PNK phosphorylation allow the structure to be tracked by agarose gels.

B.2. Results

To make the forked and spanning structures, we used the pBC-SK(-) and pBC-SK(+) phagemids. Once transformed into E. coli containing the F' episome, these phagemids can be infected with a helper phage to produce circular single-stranded DNA molecules that are complementary in all regions except the f1 origin (Figure B.2b). We envisioned that the forked and spanning structures could be completed by annealing contiguous primers that are complementary to the f1 origins (in a forked or spanning arrangement) followed by ligation of the primers in vitro (Figure B.2c). However, this method did not produce the desired products with high efficiency (Figure B.3), possibly due to primers that are incomplete and non-contiguous.

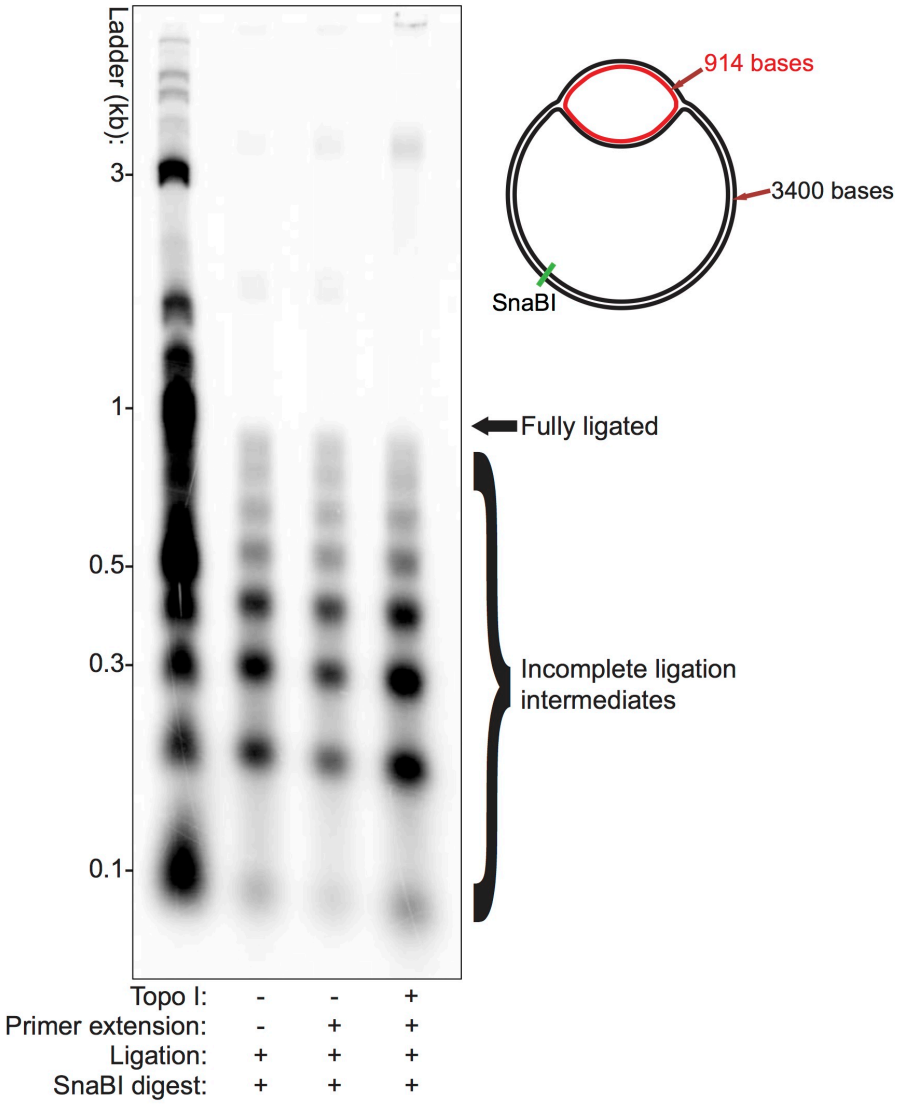


Figure B.3: Construction of the spanning DNA structure in vitro is not efficient. Denaturing agarose gel tracking 32P-ATP-labeled primers. The desired fully-ligated 914bp inner circle is a minor product of ligation. A primer extension reaction designed to extend incomplete primers or a TopoI treatment aimed at relieving any structural stress do not improve the construction efficiency.

Given the range of biochemical activity in frog egg extracts, we tested the ability of HSS to complete the spanning and forked structures. In fact, incubating the annealed spanning structure components in HSS generates a spanning structure within ~3 minutes (Figure B.4). More importantly, this structure is stable for at least 20 minutes, allowing for the licensing of replication origins.

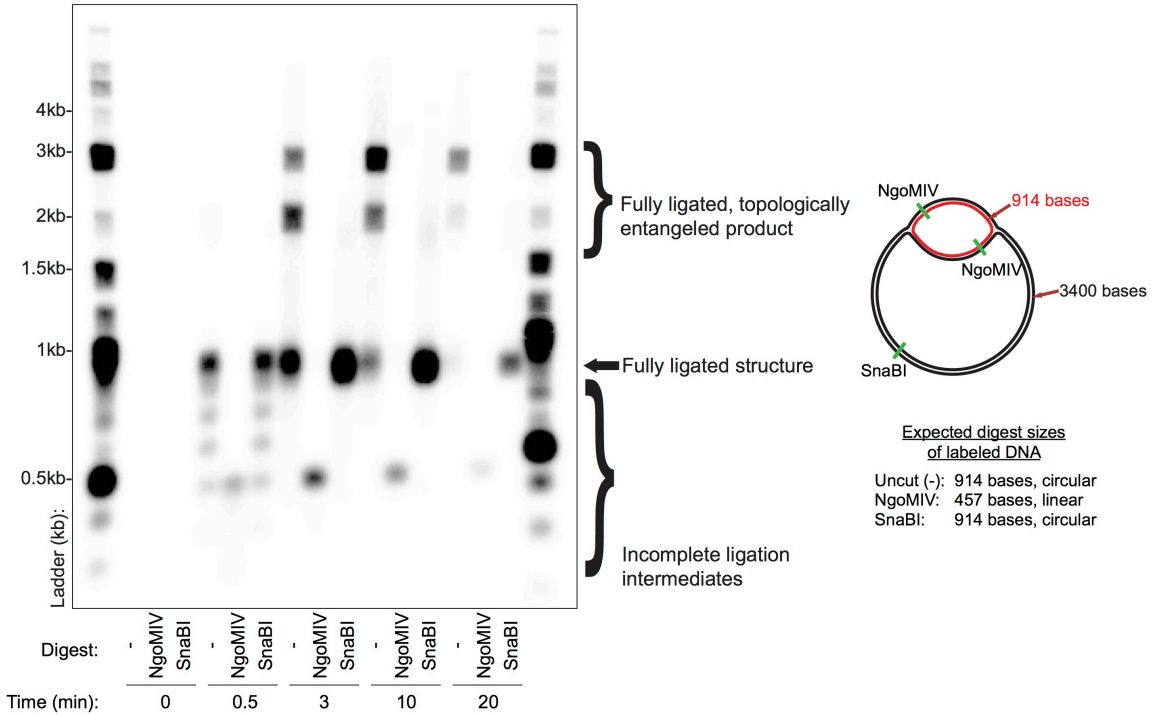


Figure B.4: The spanning DNA structure is completed after incubation in HSS. Denaturing agarose gel of 32P-ATP labeled spanning structure. Incubating the annealed DNA components of the spanning structure in HSS produces the full spanning DNA structure. While the primers in this assay have been 5' phosphorylated with radioactive ATP, the in vitro phosphorylation step is not necessary for efficient structure formation (data not shown).

We then tested the ability of HSS to complete the forked DNA structure. The desired structure is also formed within ~3 minutes, but then degenerates into a number of high molecular weight products by 10 minutes (Figure B.5). The size of products formed from incubation of the forked structure in HSS (Figures B.5 and B.7) agrees well with those expected from nuclease digestion of the forked structure (Figure B.6 – Groups of bands should be present at 1185-1513bbp, 2016-2344bp, and 3370-3857bp). Given the instability of the forked structure in HSS, we did not test it further in replication experiments.

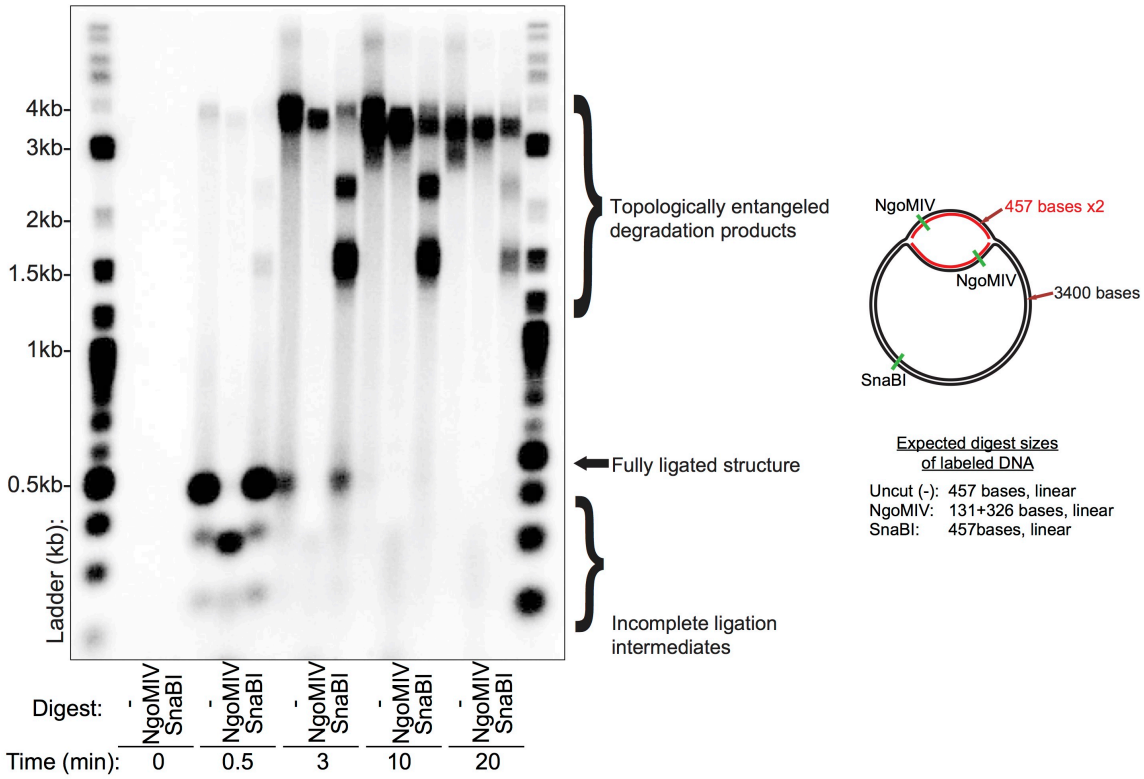


Figure B.5: The forked DNA structure is completed in HSS but then promptly disintegrates. Denaturing agarose gel of 32P labeled forked structure. Incubating the annealed DNA components of the forked structure in HSS produces the full spanning DNA structure, which then disintegrates into high molecular weight products.

To test whether the spanning structure could be replicated and to identify the products of its replication, we generated and licensed spanning structure in HSS prior to NPE addition. We tracked the replication products in two ways: 1) by supplementing the replication mix with 32P-dATP to monitor nascent DNA, and 2) by using primers that are 5' phosphorylated with 32P-ATP in order to monitor the fate of the small, inner circle.

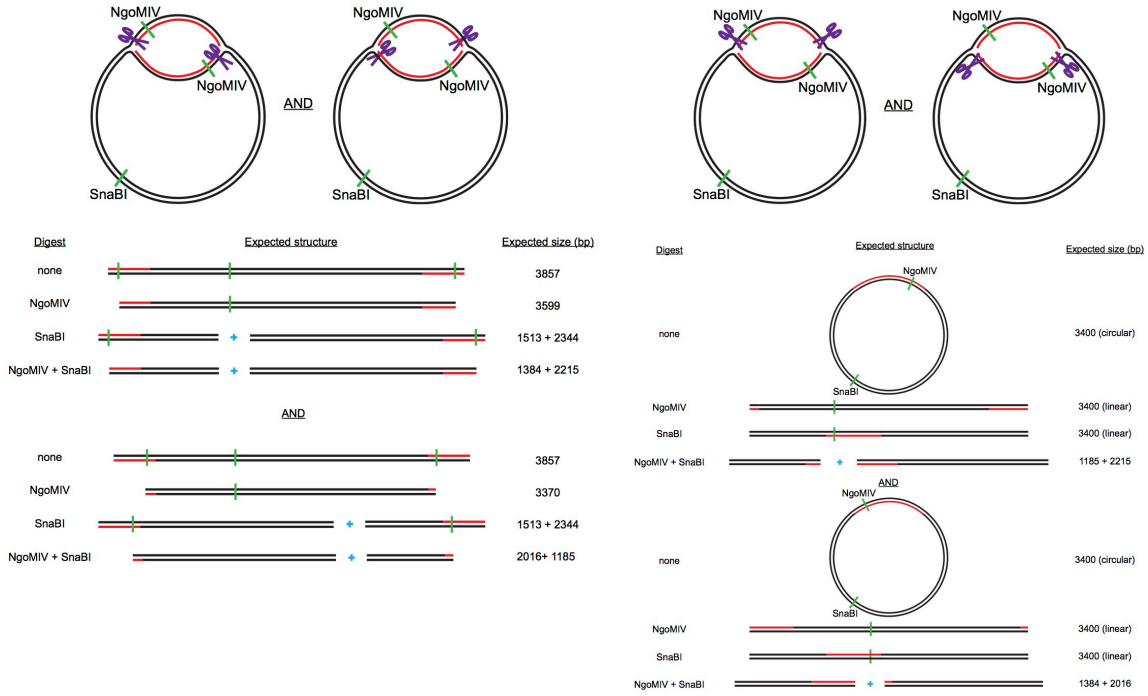


Figure B.6: Theoretical products resulting from nuclease processing of the forked DNA structure.

B.2.1. 32P-dATP-supplemented replication of spanning structure

Two major nascent products emerge after the spanning structure is subjected to replication conditions. One product is pBC-SK plasmid-sized and one product corresponds to the size of the inner, small circle of the full spanning structure (914bp) (Figures B.8, B.9). The pBC-SK-sized product could emerge from replication initiation in the 2.9kb pBC-SK(+) and (-) complementary region of the spanning structure, or from replication initiation inside the 914 small circle and completion by leading strand synthesis.

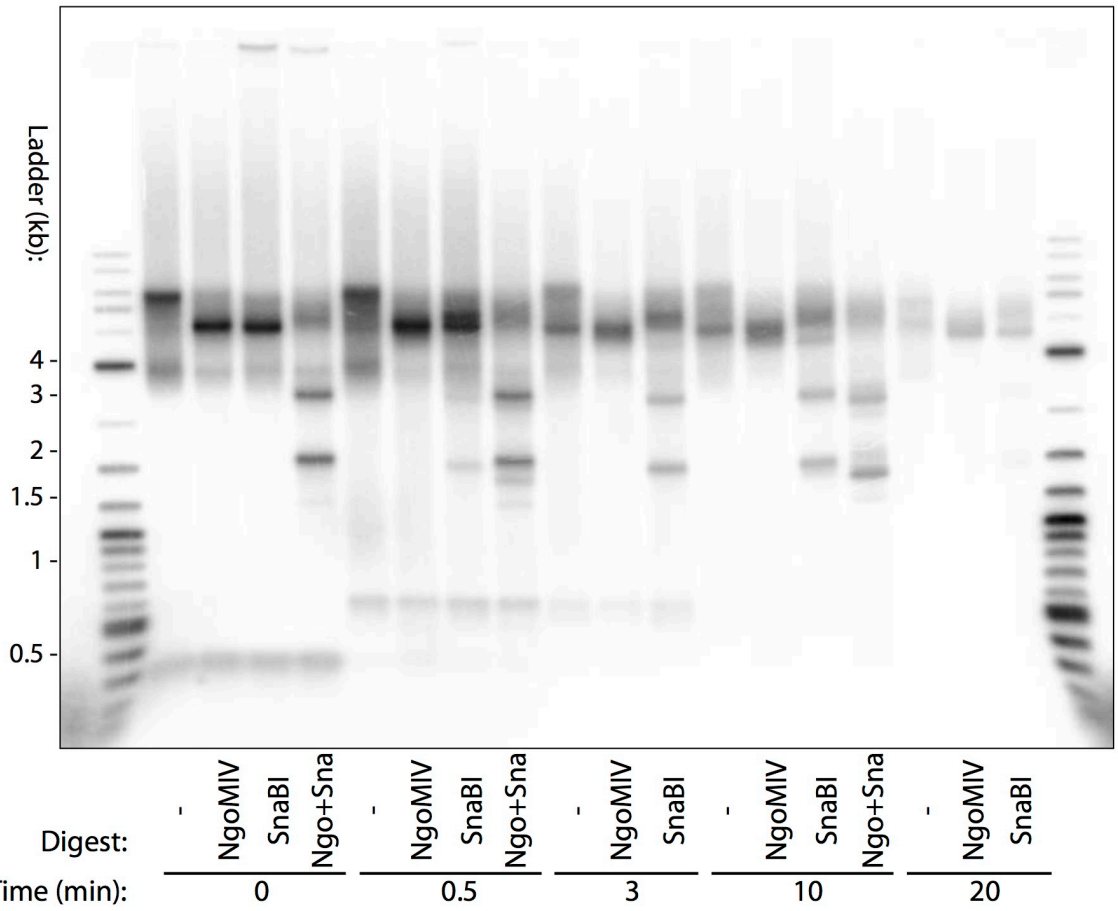


Figure B.7: The forked DNA structure is completed in HSS but then promptly disintegrates. Native agarose gel of 32P-ATP labeled forked structure.

The smaller product of the spanning structure replication could emerge from replication initiation inside the 914 small circle and completion by leading strand synthesis or initiation in the 2.9 kb region of the spanning structure, extrusion of the small single-stranded inner circle, and subsequent fill-in synthesis. The latter scenario is part of the proposed mechanism for inverted repeat amplification. Since circular ssDNA is turned into dsDNA in frog egg extracts (personal communication, Walter lab), the end product for both scenarios is a double-stranded ~1kb piece of DNA, which we see in our experiments.

B.2.2. Replication of 32P-ATP-labeled spanning structure

The major products that emerge from replication of 32P-ATP-labeled spanning structure have an approximate size of ~1kb and are similar in migration in a native agarose gel to the smaller molecular weight products from 32P-dATP supplemented experiments (Figures B.8, B.10). These products are made up of a NgoMIV-sensitive population, which would correspond to a double-stranded inner circle, as well as a NgoMIV-insensitive population whose identity is unknown (Figure B.10).

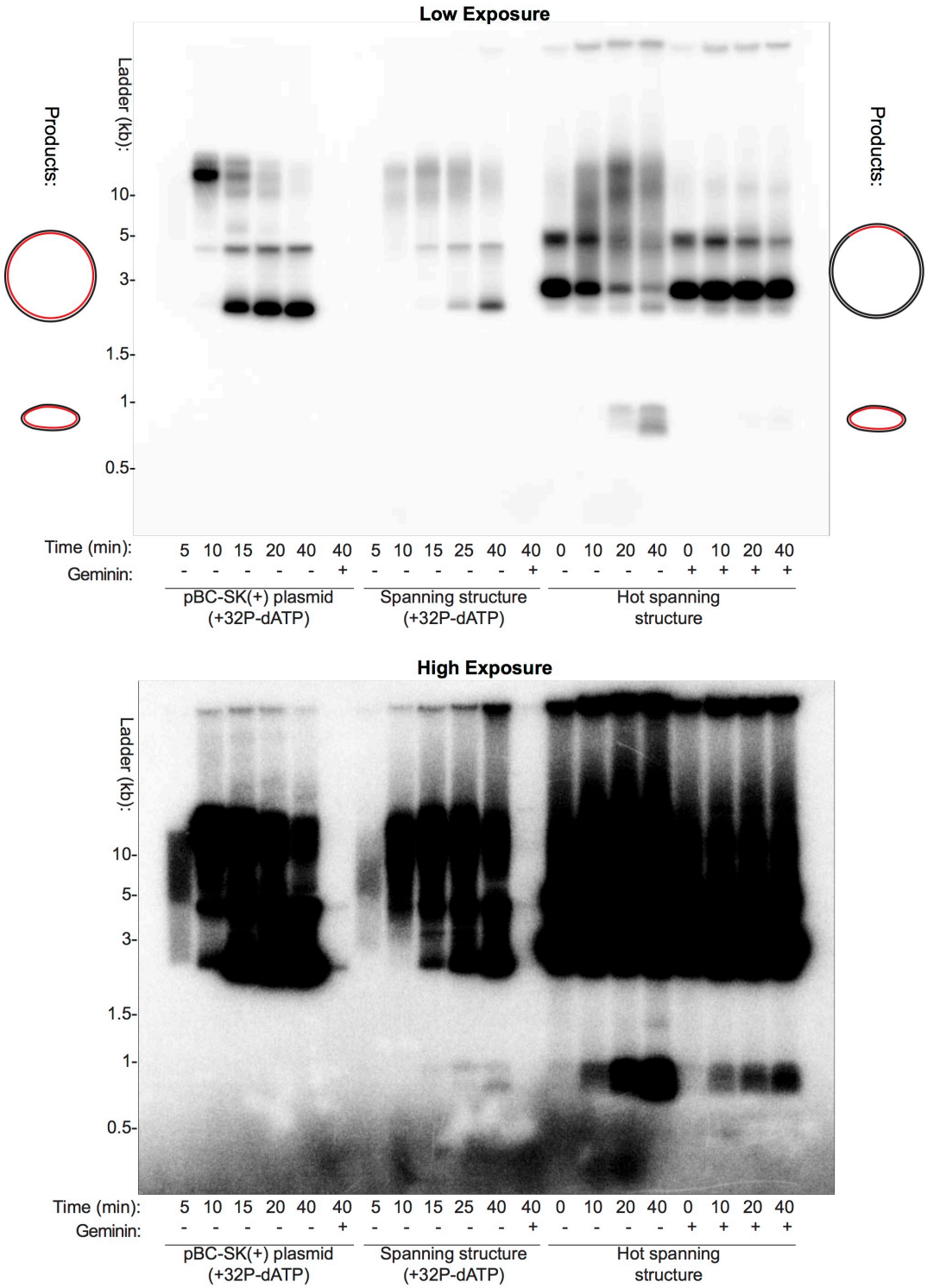


Figure B.8: The spanning DNA structure replicates to produce plasmid sized and bubble sized products.
Native agarose gel shown at two exposures. 'g' = geminin (replication inhibitor). 'hot' = 32P-ATP labeled structure (red color in 'products').

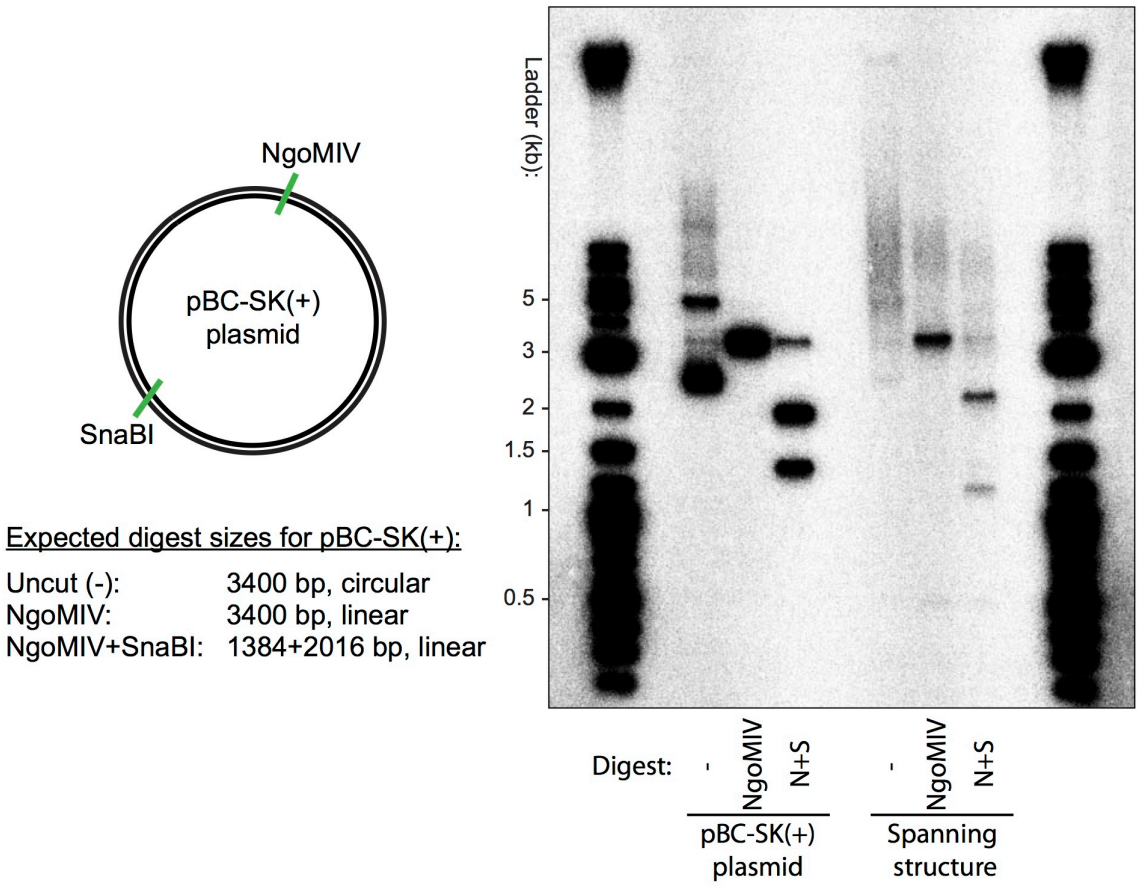


Figure B.9: Replication of the spanning structure results in plasmid sized as well as ~1kb, NgoMIV sensitive product.
Native agarose gel, Time=40 min. N+S = NgoMIV+SnaBI double digest. Double digest signal for "Spanning structure" sample corresponds to that expected from pBC-SK(+) plasmid (1185+2215 bp)..

B.2.3. Replication-dependence of final products

Although the products of 32P-dATP-supplemented replication are completely dependent on pre-RC formation (Figure B.8, compare +geminin with –geminin), the ~1kb products of 32P-ATP labeled spanning structure replication are not completely dependent on origin activity. This leads us to believe that some nuclease processing of the spanning theta structure is taking place in NPE, and that this processing is replication independent. Unfortunately, the amount of

product generated in Figure B.8 is also extract dependent. Although the amount of inner circle product is significantly larger when replication is allowed (compared to the geminin-inhibited condition) in the experiment shown in Figure B.8, those levels are relatively similar for the replicate in Figure B.11, which was performed with the same protocol, but a different batch of frog egg extract. Therefore, the variability in frog egg extract preparations confounds the degree to which replication contributes to the generation of the inner circle product, and ultimately the interpretation of how replication forks deal with spanning fork DNA structures.

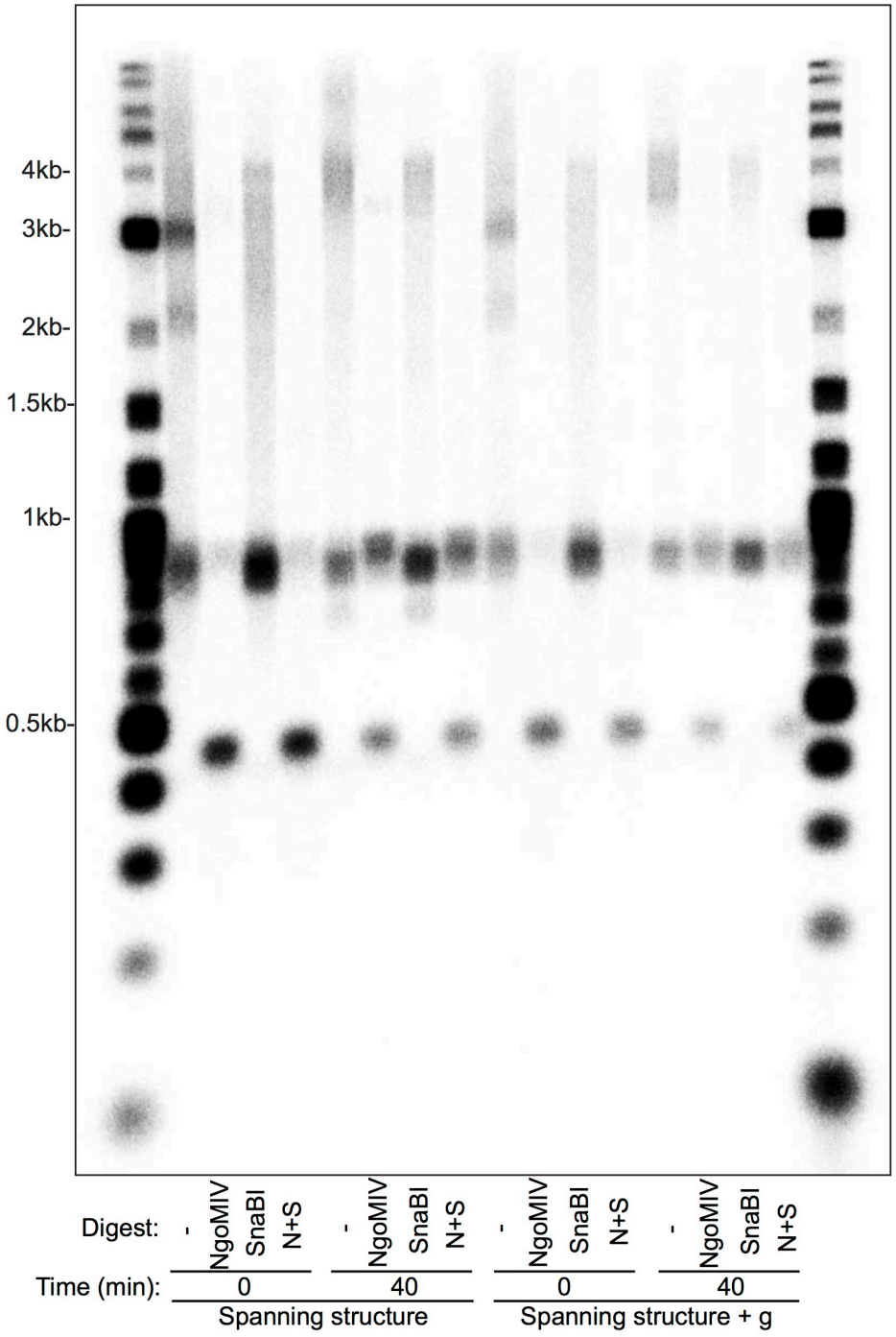


Figure B.10: Replication of the spanning structure results in NgoMIV sensitive and insensitive product.
Denaturing agarose gel. 'g' = geminin. 'N+S' = NgoMIV+SnaBI double digest.

B.2.3. Replication-dependence of final products

Although the products of 32P-dATP-supplemented replication are completely dependent on pre-RC formation (Figure B.8, compare +geminin with –geminin), the ~1kb products of 32P-ATP labeled spanning structure replication are not completely dependent on origin activity. This leads us to believe that some nuclease processing of the spanning theta structure is taking place in NPE, and that this processing is replication independent. Unfortunately, the amount of product generated in Figure B.8 is also extract dependent. Although the amount of inner circle product is significantly larger when replication is allowed (compared to the geminin-inhibited condition) in the experiment shown in Figure B.8, those levels are relatively similar for the replicate in Figure B.11, which was performed with the same protocol, but a different batch of frog egg extract. Therefore, the variability in frog egg extract preparations confounds the degree to which replication contributes to the generation of the inner circle product, and ultimately the interpretation of how replication forks deal with spanning fork DNA structures.

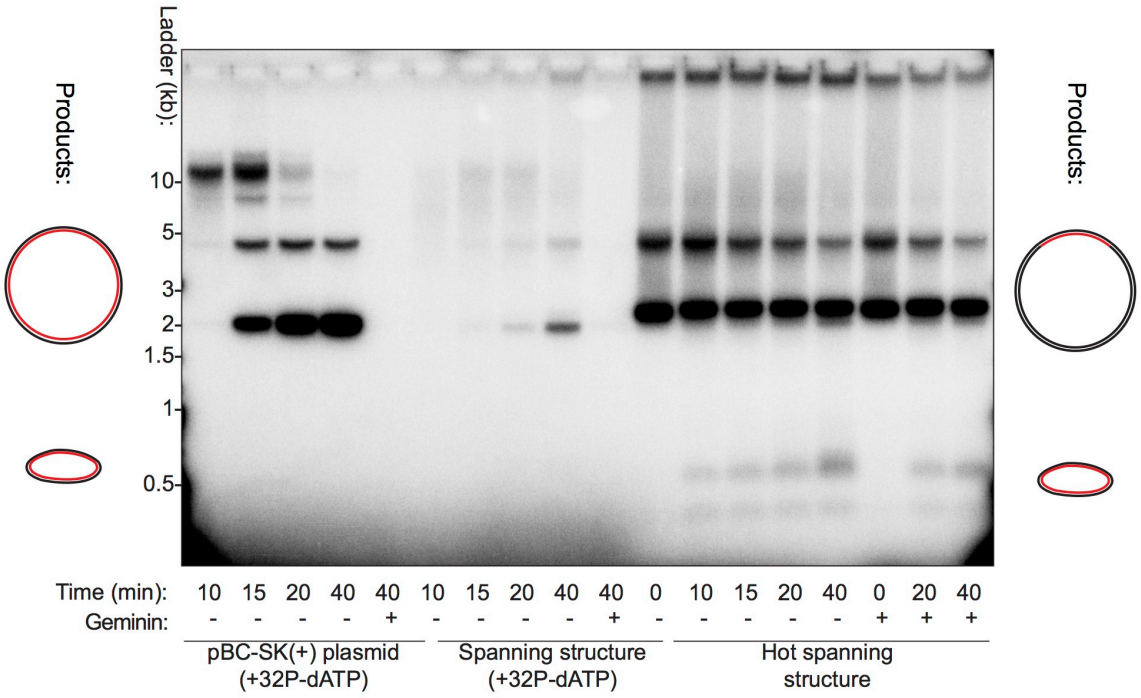


Figure B.11: Replication-dependent generation of inner circle is not reproducible from extract to extract.
The amount of product generated using the hot spanning structure at 20 and 40 minutes is similar regardless of whether replication conditions were allowed. Agarose gel. 'g' = geminin, 'N+S' = NgoMIV+SnaBI double digest..

B.3. Discussion and Future Directions

The array of structures encountered by replication forks during genomic replication spans well beyond simple double-DNA. Depending on the nature of the modified landscape and genetic background, a replication fork may become inactive and form unusual structures that need to be resolved in order for genome duplication to complete successfully. In this project, I aimed to investigate the consequences of active replication forks encountering a forked DNA structure and a spanning DNA structure, both of which have been proposed to exist in vivo (Brewer et al. 2015; Cortez 2015).

In these studies, I found that the spanning DNA structure is efficiently generated in HSS extract and remains stable. When subjected to replication tests, this structure produces a replication-dependent as well as a replication-independent ~1kb product, pointing to nuclease processing of the spanning structure. Although the replication-dependent ~1kb product can be formed from initiation inside the small spanning structure bubble, we cannot exclude that it is a product of inner circle extrusion following initiation in the 2.9kb complementary region of the pBC ssDNAs. The latter scenario is supported by the presence of a restriction enzyme-insensitive population, which may indicate an extruded single-stranded circle of DNA as proposed in the model for inverted-repeat generation.

In contrast to the spanning DNA structure, the forked structure is not stable in frog egg extracts. This instability may point to the forked structure as a substrate for processing by structure-specific nucleases. Such nucleases have been implicated in the processing of stalled and collapsed forks in various studies looking at different kinds of replication stresses (Cotta-Ramusino et al. 2005; Hu et al. 2012; Dehe et al. 2013; Kim et al. 2017). It is known that the stalling of replication forks initiates a checkpoint cascade that ultimately contributes to the stability of the fork (Branzei & Foiani 2010). However, the forked structure that is introduced into HSS extracts is comprised of naked DNA that is devoid of the signaling and protective proteins that are involved in the stalled fork response, which may ultimately lead to its instability. Given the biochemical richness of frog egg extracts, it is entirely possible that nucleases involved in the DNA damage

response recognize and process the forked structure. Possible candidates include EXO1, DNA2, and EEPD1(Cotta-Ramusino et al. 2005; Hu et al. 2012; Kim et al. 2017). The role of each of these factors may be addressed directly through immunodepletion experiments to test if the removal of each or all of them stabilizes forked structure in HSS. However, such experiments were beyond the scope of this project due to technical reasons.

As with PhADE experiments, the variability of NPE extracts affected the interpretation of results regarding the replication of the spanning structure. Although in most cases the difference is in the kinetics of the process being studied and may be disregarded, in these particular experiments the biochemical activity between replicating and non-replicating extracts confounded the results. As with PhADE, producing and using large amounts of high quality extracts may alleviate the lack of reproducibility for some of the experiments stated above.

B.4. Materials and Methods

Xenopus egg extracts preparation

Xenopus egg extract preparation was conducted as previously published (Lebofsky et al. 2009).

ssDNA purification

ssDNA from pBC-SK(-) and (+) were prepared largely as per the manufacturer's instruction, with some changes (Agilent). Briefly, each plasmid was transformed into XL1-Blue competent cells. A single colony was inoculated into 5 mls of 2XYT+Chloramphenicol (25ug/ml final, from 50mg/ml stock in EtOH), to which 5ul of M13KO7 helper phage (108 pfu/ml final,NEB) was added. The culture was grown at 37°C for ~20hrs, then aliquoted into 1.5ml tubes and centrifuged at 10k x G for 5 minutes. From each aliquot, 1 ml of supernatant was transferred into a new tube combined with 150 ul 20%PEG8000, 2.5M NaCl. The samples were vortexed and phage particles were allowed to precipitate at 4°C for 30 minutes. The tubes were then centrifuged for 5 minutes at 16k x G and the supernatant was removed. They were centrifuged again for 15 seconds and any remaining supernatant was removed. The pellet was resuspended in 400 ul resuspension buffer (0.3M NaOAc, 1mM EDTA) and vortexed vigorously. The samples were then extracted with 400 ul phenol-chloroform (Fisher) and centrifuged for 2 minutes at 16k x G. The aqueous phase was transferred to a new tube, combined with 1 ml 100% ethanol, and incubated on ice for 30 min. The tubes were then centrifuged for 5 minutes at 16k x G, the supernatant was removed, and the DNA was allowed to dry. The ssDNA does not form a single pellet at the bottom of the tube. Instead it sticks throughout the side, forming a thin, irregular film. ssDNA was then resuspended in 15 ul water and the product from each 1.5ml original prep was combined into one tube.

Spanning and forked structure construction

Two kinds of structures were constructed: Structures where the 5' ends of the bubble-filling primers were phosphorylated with gamma-32P-ATP (hot structures) and structures where the 5' ends were phosphorylated using non-radioactive ATP (cold structures). Phosphorylation was performed as per NEB instructions (T4 Polynucleotide Kinase, NEB) for cold structures, whereas 3ul of gamma-32P-ATP (3000Ci/mmol, 10mCi/mL, PerkinElmer) was used for hot structures instead of non-radioactive ATP. In all cases, primers were ordered PAGE purified, to remove incomplete synthesis products.

For each structure, the following molar mix was made: 2 : 1 : 1 (each phosphorylated primer(4 total) : pBC-SK(-) : pBC-SK(+)). The reaction was supplemented with Cutsmart buffer to a final concentration of 1x, heated to 95°C for 5 minutes, then allowed to cool down to 25°C at a rate of 1°C/minute in a thermocycler.

For in vitro ligation and structure completion, manufacturer's instructions were used (T4 DNA Ligase, NEB). To complete the structures in HSS, equal volumes of annealed structure and activated HSS were combined with for 20 minutes at room temperature. The reaction was then stopped by the addition of an equal volume of Stop Solution (Lebofsky et al. 2009). The completed structures were purified using columns and eluted in water at half the volume of the original structure (DNA Clean and Concentrator, Zymo).

DNA replication experiments

Replication experiments using cold structures were performed as originally published (Lebofsky et al. 2009). The same was done for replication experiments using hot structures, except using water instead of alpha-32P-dATP.

In both cases, specific time points were taken when indicated in each experiment and processed as originally published.

Restriction enzyme digests and agarose gels

All restriction enzyme reactions were performed as per manufacturer's instructions (NEB).

Replication experiments were resolved using 0.8% agarose gels in TBE. Gels were placed between pieces of DEAE paper, initially dried using paper towels dried, then vacuum dried before being exposed to Fujifilm BAS phosphorimagers before being scanned using a Typhoon machine.

For denaturing conditions, replication experiments were loaded on 1.5% denaturing agarose gels and run at 1.5V/cm under denaturing conditions (50mM NaOH, 1mM EDTA). The gel was then soaked in 7% trichloroacetic acid before being dried and visualized using a phosphorimager.

APPENDIX C: Single-Molecule Counting of MCM Hexamers on Origins of Replication

C.1. Introduction

As introduced in Chapter I, in vitro and in vivo experiments indicate that MCMs can be loaded on origins of replication in excess of the minimal double-hexamer needed for replication (Ticau et al. 2015; Das & Rhind 2016). Data also indicates that one factor involved in determining the replication timing program in budding yeast is the number MCM helicases loaded at origins. Origins that have more MCMs tend to fire earlier in S phase than the origins with fewer MCMs loaded (Das & Rhind 2016). While bulk in vivo assays suggest a correlation between origin firing and the number of helicases loaded, they do not shed light on the distribution of MCMs at individual origins or the mechanism of multiple MCM loading. Many questions remain: What is the distribution of MCM loading on early versus late firing origins? How do mutations to various origin components, such as B2 element and ACS, affect the number of MCMs loaded? What can MCM loading distributions tell us about the mechanism of MCM loading?

Methods such as ChIP-seq are very useful in relaying information about the relative levels MCM bound at multiple origin locations (Chapter II). However, this method is unable to reveal the absolute number of bound molecules. Western blot measurements can provide information about stoichiometry if combined with well-characterized protein-DNA systems. For example, by modifying a plasmid with a zinc finger binding site that specifically and tightly binds only one tagged zinc-finger protein, the number of similarly-tagged MCM loaded on that plasmid

can be estimated (Das & Rhind 2016). Though this method is useful in obtaining estimates of protein stoichiometries, it does so indirectly and may be distorted by differences in the purification efficiencies of the protein of interest and the stoichiometrically-bound protein.

One method to directly test how many MCMs are loaded on yeast origins of replication is by single-molecule microscopy counting experiments. Using a variety of microscopy setups, this methods has been applied to a number of different questions and biological systems, including determining the number of membrane bound proteins in Xenopus oocytes, the stoichiometry of packaging RNA ring in bacteriophage phi29, the number of ORC, Cdt1, and geminin molecules bound to ORCA in human cells, and the binding stoichiometries E. coli UvrB helicase (Shu et al. 2007; Ulbrich & Isacoff 2007; Shen et al. 2012; Yokota et al. 2013). Though the microscopy setup may differ based on the model organism and specific question, these assays depend on the same physical phenomenon: when organic dyes and fluorescent proteins attached to a protein of interest photobleach, their fluorescence shows irreversible and quantifiable drops in intensity. The number of photobleaching steps depends on the number of fluorophores present in a diffraction-limited area, thus making it possible to count the number of molecules present. In order to increase signal to noise ratio, photobleaching conditions can be combined with total-internal fluorescence microscopy (TIRF), which allows for the illumination of a flow cell at only ~100nm

depth in order to restrict the number of fluorophores that are excited and to reduce background noise.

Using the principles discussed above, we designed an assay where plasmids carrying specific origins of replication can be tethered to microscope slides for accurate counting of fluorescently-tagged loaded MCM hexamers by single-molecule photobleaching. Though the experiments would be performed on purified extracts, the loading of origins of replication would take place in vivo, providing direct evidence regarding the absolute number of MCM hexamers loaded on single origins of replication.

C.2. Results

In order to directly measure the number of MCMs loaded on single origins of replication, we employed the TALO8 minichromosome system. TALO8 contains a 1.45kb region of the genome that includes the gene Trp1 and the efficient origin of replication ARS1. Importantly, TALO8 contains an array made up 8 tandem LacO sequences that allows for efficient pulldown using immobilized LacI protein (Unnikrishnan et al. 2010).

We envisioned combining the ability to pull down the origin-containing TALO8 with the SNAP/CLIP-tag labeling technologies. The SNAP tag is a 20kd tag encoding a mutated version of an alkyltransferase involved in DNA repair, which acts specifically with benzylguanine (BG) groups to form irreversible covalent

bonds (Keppler et al. 2003). The BG group can be functionalized with various groups, including biotin and fluorescent dyes, allowing for specific labeling of SNAP-tagged proteins. The CLIP tag, on the other hand, is a version of the SNAP tag whose specificity was modified to react with benzylcytosine (BC) groups (Gautier et al. 2008). Like BG, BC groups can also be functionalized with an array of fluorophores and probes.

In order to count the number of MCMs loaded on TALO8 origins, we decided to tag MCM4 with a CLIP tag at its C terminus (Figure C1.a.). C-terminal tagging of MCM4 with even larger tags has been used in numerous studies and does not affect MCM4 function (Labib et al. 1999; Nguyen et al. 2000; Wu et al. 2012). In the same strain, we inserted a copy of the LacI gene tagged at its C-terminus with a SNAP tag (Figure C1.a.). Lastly, the strain was transformed with the TALO8 minichromosome, which is stably propagated at ~50 copies/cell (Unnikrishnan et al. 2010).

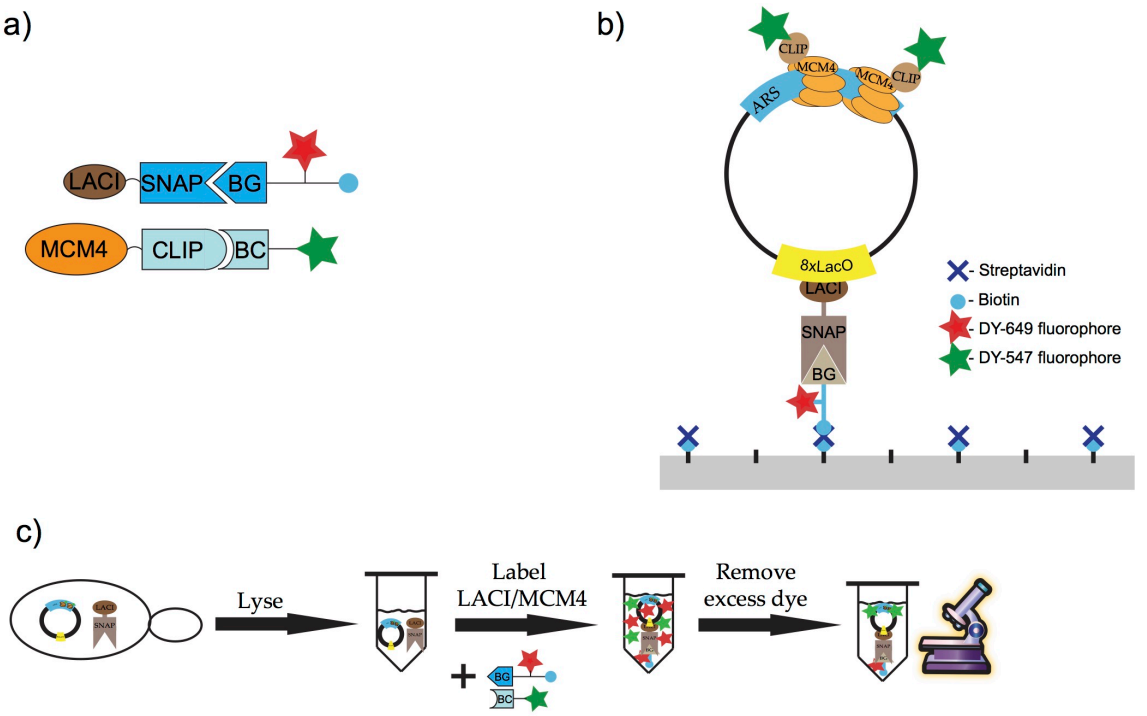


Figure C.1: Counting MCM hexamers on origins of replication using single-molcule photobleaching.
a) Orthogonal labeling of LACI-SNAP fusion protein with benzylguanine-biotin bifunctional dye and MCM4-CLIP fusion protein with benzylcytosine-conjugated dye. b) Diagram of TALO8 plasmid with labeled and loaded MCM hexamers tethered to a surface. c) Experiment scheme for purification and visualization of labeled TALO8 plasmid complexes..

Cells harboring the genetic modifications described above were arrested in G1 and harvested in order to prepare extracts for microscopy. Yeast lysates were then subjected to labeling reactions. MCM4-CLIP was labeled using BC-547, while LACI-SNAP was labeled using a bifunctional BG-biotin-649 dye (Figure C1.b.). The latter was crucial to our experimental setup as it allows for the colocalization of TALO8-bound LACI with TALO-loaded MCM4 as an additional step to ensure that non-specifically bound MCM4 does not confound the data. As seen in Figure C2.a, MCM4-CLIP and LACI-SNAP can be efficiently labeled with their respective dyes without any cross-reactivity.

Before being analyzed by TIRF microscopy, yeast extracts were subjected to a variety of purification methods to remove unreacted dyes. These methods included gravity filtration through G25 Sephadex column, Superdex 200 gel filtration, Superose 6 gel filtration, and centrifugal concentrators of various molecular weight cutoff. Of these methods, only Superose 6 gel filtration was able to efficiently separate unreacted dye from labeled proteins. Using this method, unreacted dye was efficiently separated from plasmid containing fractions during purification (Figures C2.b and C2.c). Due to the expected low abundance of plasmid-loaded MCM4-CLIP relative to total MCM4, labeled protein from purified fractions was not expected to be seen on SDS-PAGE gels or dot blots. Lastly, TALO8 purification could be tracked by sampling fractions for the presence of the TALO8 DNA using quantitative PCR (Figure C2.b).

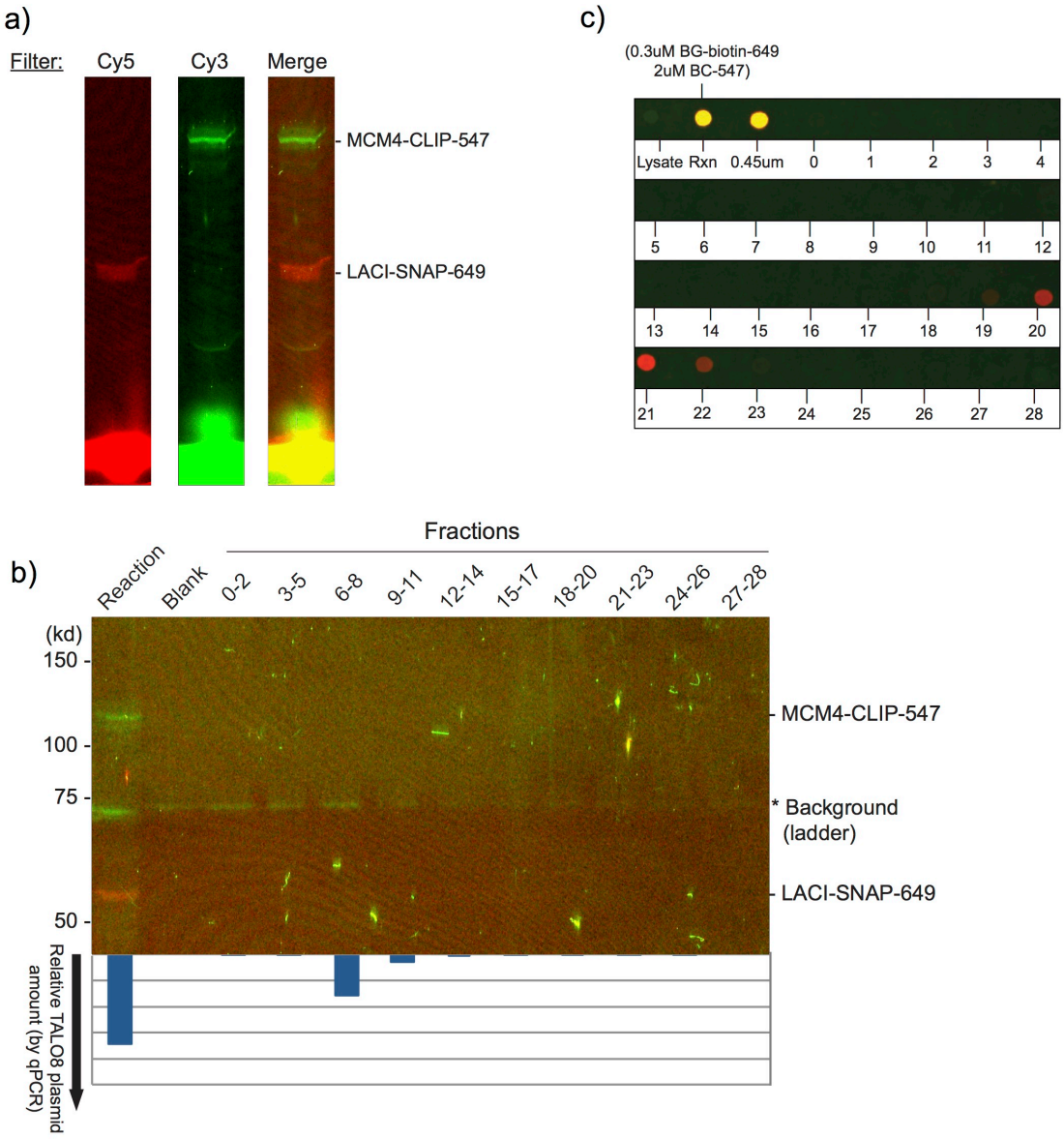


Figure C.2. Labeling and purification of yeast extracts using a Superose 6 c
a) MCM4-CLIP and LACI-SNAP can orthogonally labeled with organic dyes. 2) TAL08 minichromosomes can be purified from yeast extracts by using Superose 6 column chromatography. Fractions from the purification were run on a SDS-PAGE gel and imaged by fluorescence (Cy3/Cy5 filter merge) and tested for the presence of TALO8 by qPCR. c) Dot blots show that unreacted dye (fractions 20-22) are efficiently separated from plasmid containing fractions (6-8) (Cy3/Cy5 filter merge).

Samples containing TALO8 plasmid by qPCR were analyzed by single-molecule TIRF microscopy (fractions 6-8 in Figure C2.b). To test whether LACI-SNAP-biotin-649 binding to slides was biotin-dependent, slides were either

functionalized with streptavidin or not. As seen in Figure C3.a., only slides that were functionalized with streptavidin showed single-molecule signal, indicating that LACI-SNAP-biotin-649 binds specifically to streptavidin. In addition, subjecting the diffraction-limited spots to photobleaching conditions indicated that the signal came from single molecules of 649 dye and therefore single LACI-SNAP-biotin-649 molecules that we expected to be bound to TALO8 plasmid (Figure C3.b.).

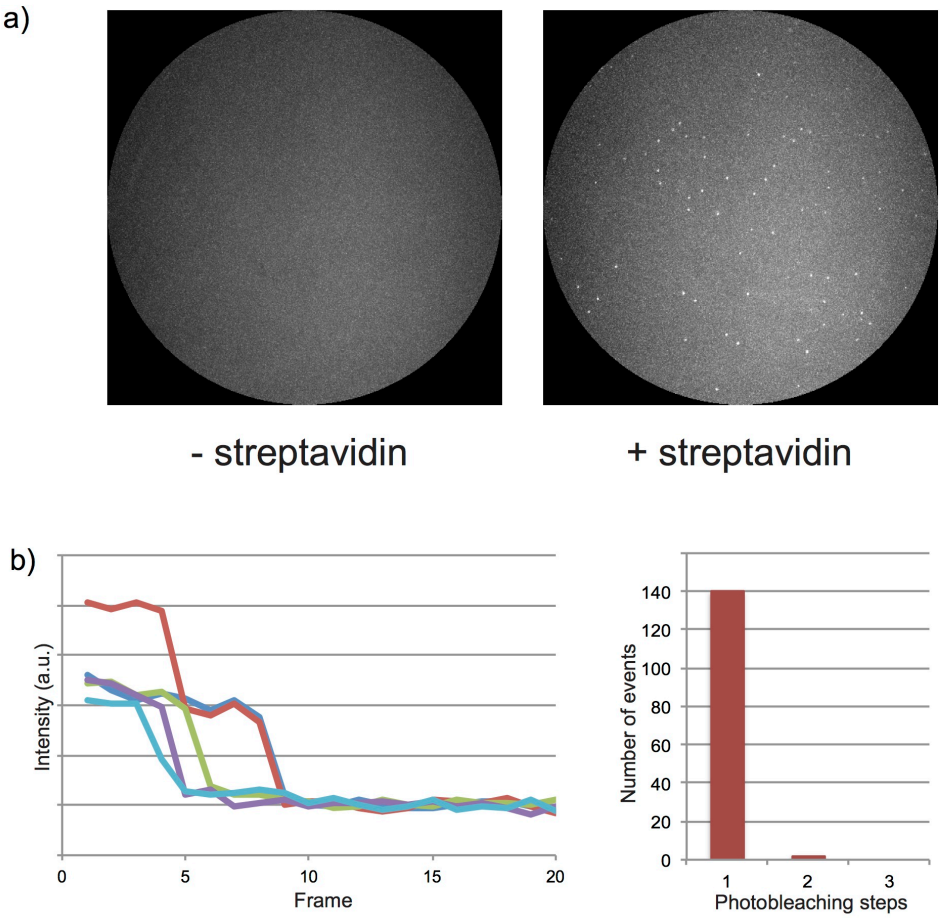


Figure C.3. Bifunctional dye BG-biotin-649 binds specifcially to the flow cell surface. a) Superose 6-purified extracts were flown into cells that were or were not functionalized with streptavidin. Single-molecule fluorescence was only observed in streptavidin-functionalized flow cells. b) Single-molecules that were subjected to photobleaching conditions generally and showed only 1 step decrease in fluorescence.

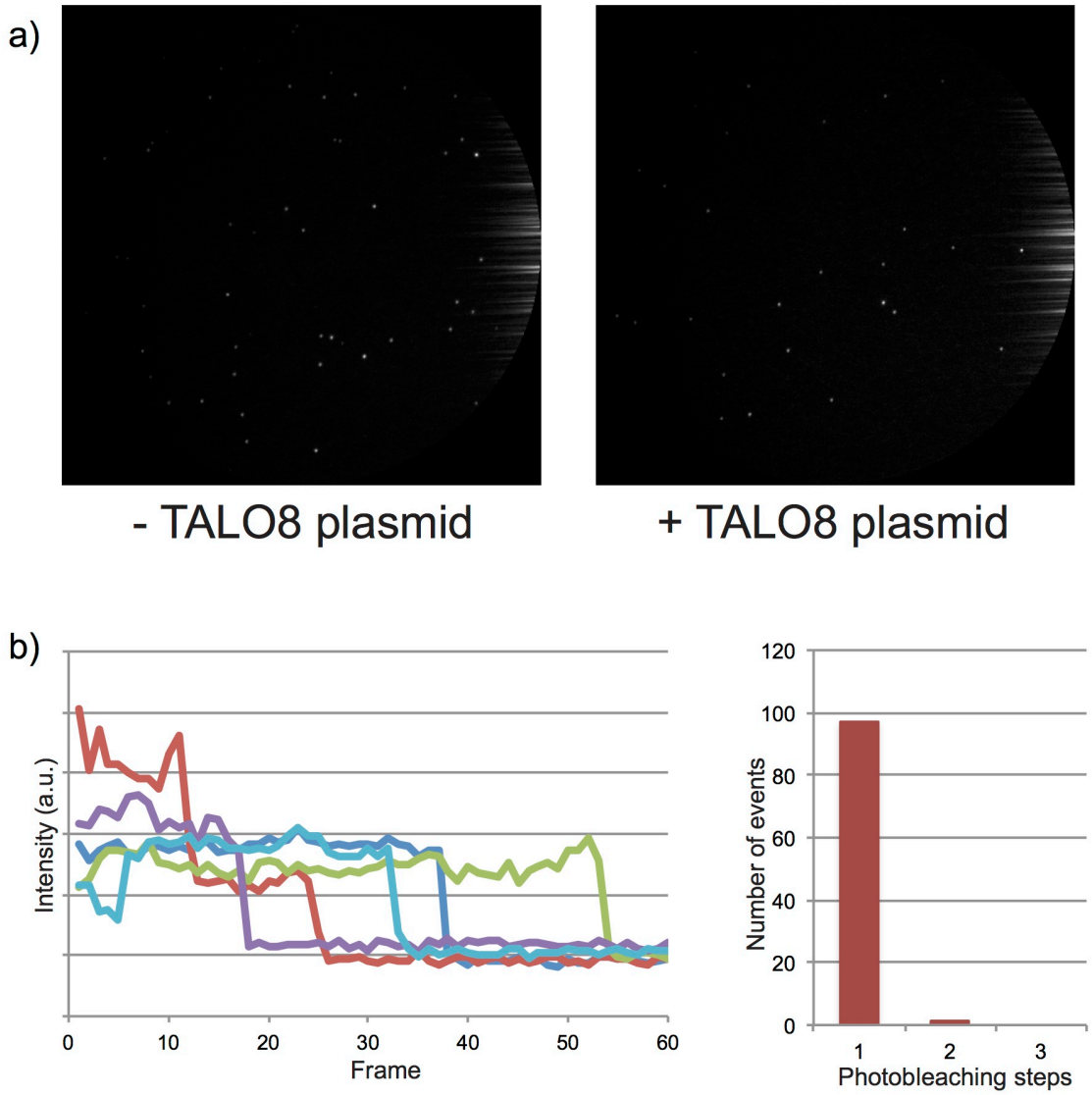


Figure C.4. The dye BC-547 binds non-specifcially to the flow cell surface.
a) Superose 6-purified extracts containing or not containing TALO8 plasmids were flown into streptavidin-functionalized flow cells. Single molecules of BC-547 were observed regardless of the presence of TALO8 plasmid. b) Single-molecules that were subjected to photobleaching conditions generally and showed only 1 step decrease in fluorescence.

To ensure that MCM4-CLIP-547 signal was emanating specifically from loaded hexamers and not non-specific binding of free MCM4-CLIP-547 to slides, TALO8-containing fractions were compared to fractions that did not contain any TALO8 signal by qPCR (Figure C4.a). Unfortunately, both extracts displayed 547

dye signal, indicating that these molecules were non-specifically binding to coverslips. When quantified by photobleaching, the vast majority of diffraction-limited spots indicated that they were made up of only one molecules of dye (Figure C4.b). Overall, these results suggest that MCM4-CLIP-547 binds to the coverslip in a non-specific manner. Since specific binding needs to occur in order to confidently count the number of loaded MCMs at origins of replication, this step needs to be optimized, possibly by changing the slide passivation and functionalization procedures.

C.3. Discussion and Future Directions

Genome-wide population studies have been instrumental in relaying information about how processes such as DNA replication work. However, they are often incapable and sometimes obscure the molecular details of how some processes occur. In order to gain a better understanding of how MCM loading on origins of replication in budding yeast occurs, we deigned an assay to measure the specific number of helicases loaded at ARS1 using single-molecule TIRF microscopy.

This assay is versatile in that the origin-containing part of the plasmid can be modified to test numerous scenarios. For example, ARS1 can be substituted with the late-firing ARS316. This experiment would further test the model where the number of MCMs loaded at origins of replication determines the likelihood of the origin firing and therefore its firing time during S phase. In this case, ARS316

would be expected to have less MCM4 molecules stably loaded than ARS1, as has previously been observed by western blot (Das & Rhind 2016).

An additional question to answer using this assay would be to mutate the B1 and B2 elements of ARS1, which have been shown to be important for ORC and MCM loading, respectively (Zou & Stillman 2000). While previous data has shown that MCM levels at both mutant origins would go down compared to wild type, this data would offer quantitative and direct proof of the role of these sequences in MCM loading.

Furthermore, additional early and late origins may be tested for the number of MCM that are loaded onto them to obtain a more thorough picture of whether MCM loading affects replication globally or only for a subset of origins. Importantly, these experiments may shed light on the mechanism of MCM loading at origins of replication. Specifically, they may be able to distinguish between a model where the number of MCMs loaded at an origin is determined by the space available at that origin, or a model where the number of MCMs is determined by the rate of loading. The former scenario would manifest itself as a tightly distributed peak when comparing the number of MCMs from numerous single molecules of the same origins. The latter scenario, on the other hand, would manifest itself as a widely distributed peak.

In order to set up the assay to perform the experiments stated above, yeast stably propagating the TALO8 minichromosome that contains the efficient origin

of replication ARS1 and a LacO array were labeled with fluorophores using the SNAP/CLIP protein labeling system. Specifically. MCM4 was labeled with a 547 organic dye and LACI protein was labeled with a bifunctional reagent containing biotin and 659 dye. This scheme allows for tethering of LacO-containing TALO8 plasmid to slide surfaces and for direct counting of labeled MCM4 proteins using photobleaching (Figure C1).

Labeling of both MCM4 and LACI with their respective dyes was efficient and specific, and Superose 6 gel filtration of the yeast extract separated unreacted bifunctional reagent from the LACI-functionalized population (Figure C2). In addition, tethering of LAC1-SNAP-biotin-649 onto slides was dependent on the functionalization of the slide with streptavidin (Figure C3). However, association of labeled MCM4 protein was independent of the presence of TALO8 plasmid, indicating that this association is non-specific and origin-independent in the current labeling and purification scheme (Figure C4).

In order to specifically count the number of MCM4 proteins stably loaded on TALO8 minichromosomes tethered to slides, several experimental conditions may be optimized. First, several microscope slide preparation methods may be used to gauge the conditions that produce the least non-specific binding. This parameter can be optimized at the cleaning step with the use of various harsh commercially-available cleaning solutions, including Piranha solution, NanoStrip, Micro-90 (Cardoso Dos Santos et al. 2016; Salomon et al. 2016). In addition, the

passivation method may be optimized. BSA is often used in experiments to reduce non-specific binding of proteins in extract to slides and finding the correct buffer conditions with the correct levels of BSA may yield reduced non-specific binding.

Second, the extract labeling and purification scheme has numerous spots for optimization. The dye used for MCM4 labeling, BC-547, has numerous other alternatives, including similar dyes that are excited by different wavelengths (BC-488) or dyes with different proprietary chemistry (BC-AlexaFluor series). Furthermore, other labeling chemistries, such as the HALO tag, may prove to reduce the non-specific binding of MCM4 molecules to slides.

Third, biologically encoded fluorescent proteins may also be used to label MCM4. These proteins have been used extensively for counting of single molecules using photobleaching (Ulbrich & Isacoff 2007; Shen et al. 2012; Yu et al. 2012). Tagging MCM4 with a fluorescent protein such as GFP removes the need for an exogenous labeling reaction (as with SNAP/CLIP) and has the advantage that every MCM4 in solution is labeled. To this end, I constructed various strains harboring fluorescently tagged MCM4 constructs (Table C1).

Lastly, it may be possible that, although the vast majority of the unreacted bifunctional dye is removed, a small amount may still linger in the plasmid-containing fractions after Superose 6 purification. Given the large difference in the number of molecules of dye to plasmids, it may be possible that even small

fractions of unreacted dye are binding to streptavidin-coated slides and making it unlikely to find plasmid-containing single molecules. In order to make it easier to detect molecules that actually represent plasmid-bound MCM4, I constructed a strain that expresses a HALO-tagged Zif268 protein. Zif268 is a zinc finger protein that binds a 10-bp recognition sequence with sub-nanomolar affinity (Elrod-Erickson & Pabo 1999). In addition, I also constructed a modified TALO8 minichromosome, which includes the Zif268 binding site. Combined, this background will make it possible to label Zif268-HALO and MCM4-CLIP with different dyes and, after purification, allow for the finding of MCM-loaded plasmids by using the colocalization of the two fluorescent molecules.

C.4. Materials and Methods

Yeast growth conditions and harvesting

TALO8-containing single colonies of yeast were inoculated into 2L SC-Trp media and grown at 30°C to OD(600) ~0.6, at which point they were treated with Alpha Factor (10ug/ml final) for 2 hours. Cells were collected by centrifugation and washed twice with cold water. Yeast paste from washed pellets was loaded onto 60ml syringes and expunged into liquid nitrogen in order to make “noodles.” Noodles were stored at -80°C until lysis.

Yeast lysis

Frozen noodles were lysed using a cryogenic ball mill grinder (Retsch). Briefly, while keeping the grinding vessel in liquid nitrogen, up to 3000 ODs of yeast in noodle form were added to the chamber and the yeast was ground at 400rpm for 1.5 minutes. The process was repeated 6 times, cooling the vessel to liquid nitrogen temperature between rounds. Finally, the ground yeast was weighed and stored at -80°C until labeling.

Fluorescent dye labeling of SNAP and CLIP tags

200mg ground yeast extract powder was resuspended in 450ul Resuspension Buffer (50mM Hepes pH7.4, 150mM NaCl, 10% Glycerol, 1mM DTT, 1x Complete EDTA-free protease inhibitors) in a 1.7ml tube. The samples were centrifuged at 14k x G for 15 minutes at 4°C and 200ul of the supernatant was moved into a new tube. To this supernatant was added 300nM BG-biotin-649 (NEB, not yet commercially available) and the sample was incubated at room temperature in the dark for 20 minutes. Following this reaction, CLIP-Surface-547 (NEB) was added at 2uM final for 30 minutes at room temperature in the dark. The volume was brought to 250uL using Resuspension Buffer and the sample was filtered through a 0.45um syringe filter. Samples were checked for labeling by being run on 7.5% SDS-PAGE minigels at 100V for 1.5hrs and the signal was obtained by using the corresponding filter in a Typhoon gel imager.

Yeast extract purification by Superose 6

400mg ground yeast extract powder was labeled as above. After the second room temperature labeling reaction, the samples were spun using a TLA100.4 rotor at 45k RPM for 40 minutes. The supernatant was retrieved, the volume was brought to 800ul total, and the sample was filtered through a 0.45um filter. 600ul of the filtrate was loaded onto a Superose 6 (10/300) gel filtration column. Elution of fractions was done in Resuspension Buffer. Samples from each fraction were then run on 7.5% SDS-PAGE gels to gauge unreacted dye separation. Separately, DNA from 10ul of each fraction separately cleaned by columns (Zymo) and tested for the presence of TALO8 plasmid by qPCR (Kapa SYBR Fast Universal) using LD231 and LD232 primers.

As an alternative approach, samples shown to contain TALO8 by qPCR were concentrated using 100kd centrifugal concentrators (Amicon), per manufacturer's instructions. However, this method did not yield any improvements.

Slide preparation

Slides were prepared as previously published using a poly-L-lysine-graft-PEG-biotin method and functionalized with streptavidin (Smith et al. 2019).

TIRF microscopy

Flow cells were loaded on a custom made slide holder and mounted on an inverted microscope (Olympus Ix81) equipped with an Olympus 100x 1.49 NA oil immersion objective, three Hamamatsu EM-CCD digital cameras, and Prior Nanoscan Z and H31XYZE stage controllers. Laser excitation was generated by CellTIRF lasers at 405nm, 491nm, 568nm, and 650nm.

APPENDIX D: Yeast and Bacterial Strains Used in Thesis Work

	Strain #	Genotype
yFS	975	MATahis3-11,15leu2-3,112ura3-1trp1-1ade2-1 can1-100RAD5+TALO8pFS408(CMV-LacI-SNAP
	976	? his3-11,15leu2-3,112ura3-1trp1-1ade2-1 (?)can1-100RAD5+Mcm4-GFP(HPH)LacI-SNAP(NAT)
	977	? his3-11,15leu2-3,112ura3-1trp1-1ade2-1 (?)can1-100RAD5+Mcm4-GFP(HPH)LacI-SNAP(NAT)
	979	? his3-11,15leu2-3,112ura3-1trp1-1ade2-1 (?)can1-100RAD5+Mcm4-GFP(HPH)LacI-SNAP(NAT)pFS408pFS408
	980	? his3-11,15leu2-3,112ura3-1trp1-1ade2-1can1-100RAD5+MCM4-mNeonGreen (KAN)LacI-SNAP(NAT)(pFS454 (mNeonGreen + Kan))
	981	? his3-11,15leu2-3,112ura3-1trp1-1ade2-1can1-100RAD5+MCM4-mNeonGreen (KAN)LacI-SNAP(NAT)pFS408(ARS1. 8xLacO)
	992	a his3-11,15leu2-3,112ura3trp1-1ade2-1 can1-100GAL RAD5 ssd1-d2Mcm4-GFP(HPH)pFS408
	993	a his3-11,15leu2-3,112ura3trp1-1ade2-1can1-100GAL RAD5 ssd1-d2Mcm4-GFP(HPH)LACI-SNAP (URA3)pFS408
	994	a his3-11,15leu2-3,112ura3trp1-1ade2-1can1-100GAL RAD5 ssd1-d2LACI-SNAP (URA3)pFS408
	995	a his3-11,15leu2-3,112ura3trp1-1ade2-1 can1-100GAL RAD5 ssd1-d2MCM4-CLIP (NAT)(CLIP (NAT) from pFS488)
	996	a his3-11,15leu2-3,112ura3trp1-1ade2-1 can1-100GAL RAD5 ssd1-d2MCM4-CLIP (NAT)LACI-SNAP (URA3)(CLIP
	997	a his3-11,15leu2-3,112ura3trp1-1ade2-1 can1-100GAL RAD5 ssd1-d2MCM4-CLIP (HPH)(CLIP (HPH) from pFS489)
	998	a his3-11,15leu2-3,112ura3trp1-1ade2-1can1-100GAL RAD5 ssd1-d2MCM4-CLIP (HPH)LACI-SNAP (URA3)(CLIP
	1001	a his3-11,15leu2-3,112ura3trp1-1ade2-1can1-100GAL RAD5 ssd1-d2MCM4-CLIP (NAT)LACI-SNAP (URA3)pFS408
	1009	a leu2ura3-52trp1gal2prc1-407prb1-1122pep4-3
	1017	a leu2ura3-52 trp1MCM4-GFP (hph)prb1-1122pep4-3prc1-407gal2
	1018	a leu2ura3-52 trp1gal2prc1-407prb1-1122pep4-3(pFS487@ura3)
	1019	a leu2ura3-52trp1gal2prc1-407prb1-1122pep4-3MCM4-CLIP (HPH)(pFS487@ura3)
	1020	a his3-11,15leu2-3,112ura3-1trp1-1ade2-1can1-100bar1- deltalys2::HisGPep4::unmarked
	1021	a his3-11,15leu2-3,112ura3-1 trp1-1ade2-1can1-100bar1::HisG ; lys2::HisGpep4::unmarked gal o/e MCM2-7, Cdt1,
	1023	a leu2ura3-52 trp1gal2prc1-407prb1-1122pep4-3pFS408pFS408(pFS487@ura3)
	1024	a leu2ura3-52trp1gal2prc1-407prb1-1122pep4-3MCM4-CLIP (HPH)pFS408pFS408(pFS487@ura3)
	1027	a his3-11,15leu2-3,112ura3-1trp1-1ade2-1can1-100ADH1-OsTir1-9Myc(URA3)
	1028	a his3-11,15leu2-3,112ura3-1trp1-1ade2-1can1-100ADH1-OsTir1-9Myc(URA3)Mcm4::Mcm4-aid(kanMX)
	1029	a his3-11,15leu2-3,112ura3-1 trp1-1ade2-1can1-100GAL-OsTir1-9Myc(URA3)
	1030	a his3-11,15leu2-3,112ura3-1 trp1-1ade2-1 can1-100GAL-OsTir1-9Myc(URA3)Mcm4::Mcm4-aid(kanMX)
	1044	a his3-11,15leu2-3,112ura3-1::ADH1-osTIR1-2-9Myc (URA3)trp1-1ade2-1can1-100
	1045	a his3-11,15leu2-3,112ura3-1::ADH1-osTIR1-9Myc (URA3)trp1-1ade2-1can1-100mcm4::mcm4-IAA17(kanMX)
	1050	a his3-11,15leu2-3,112ura3-1::ADH1-osTIR1-9Myc (URA3)trp1-1ade2-1can1-100mcm4::mcm4-IAA17-GFP (hph)(GFP-Hph (from pFS270))
	1059	a his3-11,15leu2-3,112ura3-1::ADH1-osTIR1-2-9Myc (URA3)trp1-1ade2-1can1-100mcm4::mcm4-IAA17-GFP (hph)(GFP-Hph from pFS270)
	1060	a his3-11,15leu2-3,112ura3-1trp1-1ade2-1can1-100bar1- deltapep4::unmarkedlys2::HisGTrp1::pFS497(pFS496@lys2 pFS498@his3
	1061	a his3-11,15leu2-3,112ura3-1trp1-1ade2-1can1-100bar1- deltapep4::unmarkedlys2::HisGTrp1::pLD27(pFS496@lys2 pFS498@his3pLD27@trp1)
	1062	a his3-11,15leu2-3,112ura3-1::ADH1-osTIR1-2-9Myc (URA3)trp1-1 ade2-1can1-100mcm4::mcm4-IAA17(kanMX)
	1075	a his3-11,15leu2-3,112ura3-1trp1-1ade2-1can1-100bar1- deltalys2::HisGPep4::unmarked(pFS496@His3 pFS498@lys2
	1076	a his3-11,15leu2-3,112ura3-1trp1-1ade2-1 can1-100bar1- deltalys2::HisGPep4::unmarked(pFS496@His3 pFS498@lys2

yLD	162	ahis3-11,15mcm4::mcm4-IAA17-GFP (hph)
	163	aleu1-32 ura3-1::ADH1-osTIR1=9Myc (URA3)ade2-1his3 - 11,15trp1-1 ; can1-100mcm4::mcm4-IAA17-V5 (HPH)
	164	ahis3-11,15mcm4::mcm4-IAA17-GFP (hph)(GFP-Hph from pFS270)
	165	ahis3-11,15Trp1:: (pSKM003 (Gal1, 10 - MCM6, MCM7)His3:: (pSKM004 (Gal1, 10 - MCM2, Flag-MCM3)
	166	ahis3-11.15lys2::HisGTrp1::pFS497(pFS496@lys2 pFS498@his3 pFS497@trp1)
	167	ahis3-11,15lys2::HisGTrp1::yFS497(pFS496@lys2 pFS498@his3 pFS497@trp1)
	168	ahis3-11,15lys2::HisGTrp1::pLD27(pFS496@lys2 pFS498@his3 pLD27@trp1)
	169	ahis3-11,15lys2::HisGTrp1::pLD27(pFS496@lys2 pFS498@his3 pLD27@trp1)
	170	ahis3-11,15mcm4::mcm4-IAA17(kanMX)
	171	ahis3-11,15mcm4::mcm4-IAA17(kanMX)
	172	ahis3-11,15lys2::hisG(pFS498@his3)
	173	ahis3-11,15lys2::hisG(pFS498@his3 pFS496@lys2)
	174	ahis3-11,15lys2::hisG(pFS496@His3 pFS498@lys2 pLD27@trp1)
	175	ahis3-11,15lys2::HisG(pFS496@His3 pFS498@lys2 pFS497@trp1)
pLD	15	LAC!-SNAP (URA3)
	16	CLIP (NAT)
	17	CLIP (HygR)
	18	CLIP-MCM4 sgRNA plasmid
	19	Pep4 deletion sgRNA plasmid
	20	Halo tag cassette
	21	Halo tag cassette
	22	Zif268-Halo plasmid
	23	Zif268-Halo plasmid
	24	Zif268-Halo plasmid
	25	Zif268-Halo plasmid
	26	pFS408 + Zif binding site
	27	pFS497+GFP @mcm7 c-term
	28	pFS497+GFP @mcm7 c-term
	29	pFS498 clone

REFERENCES

AYYAGARI R., GOMES X.V., GORDENIN D.A. & BURGERS P.M. 2003. Okazaki fragment maturation in yeast. I. Distribution of functions between FEN1 AND DNA2. *J Biol Chem* 278: 1618-1625.

AZMI I.F., WATANABE S., MALONEY M.F., KANG S., BELSKY J.A., MACALPINE D.M., PETERSON C.L. & BELL S.P. 2017. Nucleosomes influence multiple steps during replication initiation. *Elife* 6:

BAE S.H., BAE K.H., KIM J.A. & SEO Y.S. 2001. RPA governs endonuclease switching during processing of Okazaki fragments in eukaryotes. *Nature* 412: 456-461.

BAXTER J. & DIFFLEY J.F. 2008. Topoisomerase II inactivation prevents the completion of DNA replication in budding yeast. *Mol Cell* 30: 790-802.

BELL S.P. & STILLMAN B. 1992. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* 357: 128-134.

BELSKY J.A., MACALPINE H.K., LUBELSKY Y., HARTEMINK A.J. & MACALPINE D.M. 2015. Genome-wide chromatin footprinting reveals changes in replication origin architecture induced by pre-RC assembly. *Genes Dev* 29: 212-224.

BESNARD E., BABLED A., LAPASSET L., MILHAVET O., PARRINELLO H., DANTEC C., MARIN J.M. & LEMAITRE J.M. 2012. Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat Struct Mol Biol* 19: 837-844.

BICKNELL L.S., BONGERS E.M., LEITCH A., BROWN S., SCHOOTS J., HARLEY M.E.,

AFTIMOS S., AL-AAMA J.Y., BOBER M., BROWN P.A., VAN BOKHOVEN H., DEAN J., EDREES A.Y., FEINGOLD M., FRYER A., HOEFSLOOT L.H., KAU N., KNOERS N.V., MACKENZIE J., OPITZ J.M., SARDA P., ROSS A., TEMPLE I.K., TOUTAIN A., WISE C.A., WRIGHT M. & JACKSON A.P. 2011a. Mutations in the pre-replication complex cause Meier-Gorlin syndrome. *Nat Genet* 43: 356-359.

BICKNELL L.S., WALKER S., KLINGSEISEN A., STIFF T., LEITCH A., KERZENDORFER C., MARTIN C.A., YEYATI P., AL SANNA N., BOBER M., JOHNSON D., WISE C., JACKSON A.P., O'DRISCOLL M. & JEGGO P.A. 2011b. Mutations in ORC1, encoding the largest subunit of the origin recognition complex, cause microcephalic primordial dwarfism resembling Meier-Gorlin syndrome. *Nat Genet* 43: 350-355.

BLUMENFELD B., BEN-ZIMRA M. & SIMON I. 2017. Perturbations in the Replication Program Contribute to Genomic Instability in Cancer. *Int J Mol Sci* 18:

BOWERS J.L., RANDELL J.C., CHEN S. & BELL S.P. 2004. ATP hydrolysis by ORC catalyzes reiterative Mcm2-7 assembly at a defined origin of replication. *Mol Cell* 16: 967-978.

BRANZEI D. & FOIANI M. 2010. Maintaining genome stability at the replication fork. *Nat Rev Mol Cell Biol* 11: 208-219.

BREWER B.J., PAYEN C., DI RIENZI S.C., HIGGINS M.M., ONG G., DUNHAM M.J. & RAGHURAMAN M.K. 2015. Origin-Dependent Inverted-Repeat Amplification: Tests of a Model for Inverted DNA Amplification. *PLoS Genet* 11: e1005699.

BREWER B.J., PAYEN C., RAGHURAMAN M.K. & DUNHAM M.J. 2011. Origin-

dependent inverted-repeat amplification: a replication-based model for generating palindromic amplicons. *PLoS Genet* 7: e1002016.

BROACH J.R., LI Y.Y., FELDMAN J., JAYARAM M., ABRAHAM J., NASMYTH K.A. & HICKS J.B. 1983. Localization and sequence analysis of yeast origins of DNA replication. *Cold Spring Harb Symp Quant Biol* 47 Pt 2: 1165-1173.

BURGERS P.M. 2009. Polymerase dynamics at the eukaryotic DNA replication fork. *J Biol Chem* 284: 4041-4045.

BURKHART R., SCHULTE D., HU D., MUSAHL C., GÖHRING F. & KNIPPERS R. 1995. Interactions of human nuclear proteins P1Mcm3 and P1Cdc46. *Eur J Biochem* 228: 431-438.

BYUN T.S., PACEK M., YEE M.C., WALTER J.C. & CIMPRICH K.A. 2005. Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint. *Genes Dev* 19: 1040-1052.

CADET J., MOURET S., RAVANAT J.L. & DOUKI T. 2012. Photoinduced damage to cellular DNA: direct and photosensitized reactions. *Photochem Photobiol* 88: 1048-1065.

CARDOSO DOS SANTOS M., VÉZY C. & JAFFIOL R. 2016. Nanoscale characterization of vesicle adhesion by normalized total internal reflection fluorescence microscopy. *Biochim Biophys Acta* 1858: 1244-1253.

CAYROU C., COULOMBE P., VIGNERON A., STANOJCIC S., GANIER O., PEIFFER I., RIVALS E., PUY A., LAURENT-CHABALIER S., DESPRAT R. & MÉCHALI M. 2011. Genome-scale analysis of metazoan replication origins reveals their

organization in specific but flexible sites defined by conserved features. *Genome Res* 21: 1438-1449.

CHANG F., MAY C.D., HOGGARD T., MILLER J., FOX C.A. & WEINREICH M. 2011. High-resolution analysis of four efficient yeast replication origins reveals new insights into the ORC and putative MCM binding elements. *Nucleic Acids Res* 39: 6523-6535.

CHATRATH P., SCOTT I.S., MORRIS L.S., DAVIES R.J., BIRD K., VOWLER S.L. & COLEMAN N. 2006. Immunohistochemical estimation of cell cycle phase in laryngeal neoplasia. *Br J Cancer* 95: 314-321.

CHEN C.L., RAPPAILLES A., DUQUENNE L., HUVET M., GUILBAUD G., FARINELLI L., AUDIT B., D'AUBENTON-CARAFA Y., ARNEODO A., HYRIEN O. & THERMES C. 2010. Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes. *Genome Res* 20: 447-457.

CHEN S., DE VRIES M.A. & BELL S.P. 2007. Orc6 is required for dynamic recruitment of Cdt1 during repeated Mcm2-7 loading. *Genes Dev* 21: 2897-2907.

CHILKOVA O., STENLUND P., ISOZ I., STITH C.M., GRABOWSKI P., LUNDSTRÖM E.B., BURGERS P.M. & JOHANSSON E. 2007. The eukaryotic leading and lagging strand DNA polymerases are loaded onto primer-ends via separate mechanisms but have comparable processivity in the presence of PCNA. *Nucleic Acids Res* 35: 6588-6597.

CHRISTIANO R., NAGARAJ N., FRÖHLICH F. & WALTHER T.C. 2014. Global proteome

turnover analyses of the Yeasts S. cerevisiae and S. pombe. *Cell Rep* 9: 1959-1965.

CICCIA A. & ELLEDGE S.J. 2010. The DNA damage response: making it safe to play with knives. *Mol Cell* 40: 179-204.

CORTEZ D. 2015. Preventing replication fork collapse to maintain genome integrity. *DNA Repair (Amst)* 32: 149-157.

COSTER G. & DIFFLEY J.F.X. 2017. Bidirectional eukaryotic DNA replication is established by quasi-symmetrical helicase loading. *Science* 357: 314-318.

COSTER G., FRIGOLA J., BEURON F., MORRIS E.P. & DIFFLEY J.F. 2014. Origin licensing requires ATP binding and hydrolysis by the MCM replicative helicase. *Mol Cell* 55: 666-677.

COTTA-RAMUSINO C., FACHINETTI D., LUCCA C., DOKSANI Y., LOPES M., SOGO J. & FOIANI M. 2005. Exo1 processes stalled replication forks and counteracts fork reversal in checkpoint-defective cells. *Mol Cell* 17: 153-159.

CZAJKOWSKY D.M., LIU J., HAMLIN J.L. & SHAO Z. 2008. DNA combing reveals intrinsic temporal disorder in the replication of yeast chromosome VI. *J Mol Biol* 375: 12-19.

DAS S.P., BORRMAN T., LIU V.W., YANG S.C., BECHHOEFER J. & RHIND N. 2015. Replication timing is regulated by the number of MCMs loaded at origins. *Genome Res* 25: 1886-1892.

DAS S.P. & RHIND N. 2016. How and why multiple MCMs are loaded at origins of DNA replication. *Bioessays* 38: 613-617.

DAVÉ A., COOLEY C., GARG M. & BIANCHI A. 2014. Protein phosphatase 1 recruitment by Rif1 regulates DNA replication origin firing by counteracting DDK activity. *Cell Rep* 7: 53-61.

DAVEY M.J., INDIANI C. & O'DONNELL M. 2003. Reconstitution of the Mcm2-7p heterohexamer, subunit arrangement, and ATP site architecture. *J Biol Chem* 278: 4491-4499.

DE MOURA A.P., RETKUTE R., HAWKINS M. & NIEDUSZYNSKI C.A. 2010. Mathematical modelling of whole chromosome replication. *Nucleic Acids Res* 38: 5623-5633.

DE PICCOLI G., KATOU Y., ITOH T., NAKATO R., SHIRAHIGE K. & LABIB K. 2012. Replisome stability at defective DNA replication forks is independent of S phase checkpoint kinases. *Mol Cell* 45: 696-704.

DEEGAN T.D., BAXTER J., ORTIZ BAZÁN M.Á ?., YEELES J.T.P. & LABIB K.P.M. 2019. Pif1-Family Helicases Support Fork Convergence during DNA Replication Termination in Eukaryotes. *Mol Cell* 74: 231-244.e9.

DEHE P.M., COULON S., SCAGLIONE S., SHANAHAN P., TAKEDACHI A., WOHLSCHLEGEL J.A., YATES J.R., LLORENTE B., RUSSELL P. & GAILLARD P.H. 2013. Regulation of Mus81-Eme1 Holliday junction resolvase in response to DNA damage. *Nat Struct Mol Biol* 20: 598-603.

DIFFLEY J.F. & COCKER J.H. 1992. Protein-DNA interactions at a yeast replication origin. *Nature* 357: 169-172.

DION M.F., KAPLAN T., KIM M., BURATOWSKI S., FRIEDMAN N. & RANDO O.J. 2007.

Dynamics of replication-independent histone turnover in budding yeast. *Science* 315: 1405-1408.

DONOVAN S., HARWOOD J., DRURY L.S. & DIFFLEY J.F. 1997. Cdc6p-dependent loading of Mcm proteins onto pre-replicative chromatin in budding yeast. *Proc Natl Acad Sci U S A* 94: 5611-5616.

DRURY L.S., PERKINS G. & DIFFLEY J.F. 2000. The cyclin-dependent kinase Cdc28p regulates distinct modes of Cdc6p proteolysis during the budding yeast cell cycle. *Curr Biol* 10: 231-240.

DUZDEVICH D., WARNER M.D., TICAU S., IVICA N.A., BELL S.P. & GREENE E.C. 2015. The dynamics of eukaryotic replication initiation: origin specificity, licensing, and firing at the single-molecule level. *Mol Cell* 58: 483-494.

EATON M.L., GALANI K., KANG S., BELL S.P. & MACALPINE D.M. 2010. Conserved nucleosome positioning defines replication origins. *Genes Dev* 24: 748-753.

EDWARDS M.C., TUTTER A.V., CVETIC C., GILBERT C.H., PROKHOROVA T.A. & WALTER J.C. 2002. MCM2-7 complexes bind chromatin in a distributed pattern surrounding the origin recognition complex in Xenopus egg extracts. *J Biol Chem* 277: 33049-33057.

ELROD-ERICKSON M. & PABO C.O. 1999. Binding studies with mutants of Zif268. Contribution of individual side chains to binding affinity and specificity in the Zif268 zinc finger-DNA complex. *J Biol Chem* 274: 19281-19285.

EVRIN C., CLARKE P., ZECH J., LURZ R., SUN J., UHLE S., LI H., STILLMAN B. & SPECK C. 2009. A double-hexameric MCM2-7 complex is loaded onto origin DNA

during licensing of eukaryotic DNA replication. *Proc Natl Acad Sci U S A* 106: 20240-20245.

FACHINETTI D., BERMEJO R., COCITO A., MINARDI S., KATOU Y., KANOH Y., SHIRAHIGE K., AZVOLINSKY A., ZAKIAN V.A. & FOIANI M. 2010. Replication termination at eukaryotic chromosomes is mediated by Top2 and occurs at genomic loci containing pausing elements. *Mol Cell* 39: 595-605.

FERGUSON B.M. & FANGMAN W.L. 1992. A position effect on the time of replication origin activation in yeast. *Cell* 68: 333-339.

FORSBURG S.L. 2004. Eukaryotic MCM proteins: beyond replication initiation. *Microbiol Mol Biol Rev* 68: 109-131.

FORSLUND K., SCHREIBER F., THANINTORN N. & SONNHAMMER E.L. 2011. OrthoDisease: tracking disease gene orthologs across 100 species. *Brief Bioinform* 12: 463-473.

FRANCIS L.I., RANDELL J.C., TAKARA T.J., UCHIMA L. & BELL S.P. 2009. Incorporation into the prereplicative complex activates the Mcm2-7 helicase for Cdc7-Dbf4 phosphorylation. *Genes Dev* 23: 643-654.

FRIEDEL A.M., PIKE B.L. & GASSER S.M. 2009. ATR/Mec1: coordinating fork stability and repair. *Curr Opin Cell Biol* 21: 237-244.

FRIGOLA J., REMUS D., MEHANNA A. & DIFFLEY J.F. 2013. ATPase-dependent quality control of DNA replication origin licensing. *Nature* 495: 339-343.

GAMBUS A., VAN DEURSEN F., POLYCHRONOPOULOS D., FOLTMAN M., JONES R.C., EDMONDSON R.D., CALZADA A. & LABIB K. 2009. A key role for Ctf4 in

coupling the MCM2-7 helicase to DNA polymerase alpha within the eukaryotic replisome. *EMBO J* 28: 2992-3004.

GARNER E. & COSTANZO V. 2009. Studying the DNA damage response using in vitro model systems. *DNA Repair (Amst)* 8: 1025-1037.

GAUTIER A., JUILLERAT A., HEINIS C., CORRÊA I.R., KINDERMANN M., BEAUFILS F. & JOHNSSON K. 2008. An engineered protein tag for multiprotein labeling in living cells. *Chem Biol* 15: 128-136.

GE X.Q., JACKSON D.A. & BLOW J.J. 2007. Dormant origins licensed by excess Mcm2-7 are required for human cells to survive replicative stress. *Genes Dev* 21: 3331-3341.

GEORGESCU R.E., LANGSTON L., YAO N.Y., YURIEVA O., ZHANG D., FINKELSTEIN J., AGARWAL T. & O'DONNELL M.E. 2014. Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. *Nat Struct Mol Biol* 21: 664-670.

GEORGESCU R.E., SCHAUER G.D., YAO N.Y., LANGSTON L.D., YURIEVA O., ZHANG D., FINKELSTEIN J. & O'DONNELL M.E. 2015. Reconstitution of a eukaryotic replisome reveals suppression mechanisms that define leading/lagging strand operation. *Elife* 4: e04988.

GHAEMMAGHAMI S., HUH W.K., BOWER K., HOWSON R.W., BELLE A., DEPHOURE N., O'SHEA E.K. & WEISSMAN J.S. 2003. Global analysis of protein expression in yeast. *Nature* 425: 737-741.

GIAGINIS C., VGENOPOULOU S., VIELH P. & THEOCHARIS S. 2010. MCM proteins as diagnostic and prognostic tumor markers in the clinical setting. *Histol*

Histopathol 25: 351-370.

GOFFEAU A., BARRELL B.G., BUSSEY H., DAVIS R.W., DUJON B., FELDMANN H., GALIBERT F., HOHEISEL J.D., JACQ C., JOHNSTON M., LOUIS E.J., MEWES H.W., MURAKAMI Y., PHILIPPSEN P., TETTELIN H. & OLIVER S.G. 1996. Life with 6000 genes. *Science* 274: 546, 563-7.

GONZALEZ M.A., PINDER S.E., CALLAGY G., VOWLER S.L., MORRIS L.S., BIRD K., BELL J.A., LASKEY R.A. & COLEMAN N. 2003. Minichromosome maintenance protein 2 is a strong independent prognostic marker in breast cancer. *J Clin Oncol* 21: 4306-4313.

GRANOVSKAIA M.V., JENSEN L.J., RITCHIE M.E., TOEDLING J., NING Y., BORK P., HUBER W. & STEINMETZ L.M. 2010. High-resolution transcription atlas of the mitotic cell cycle in budding yeast. *Genome Biol* 11: R24.

GUO Z., KUMAGAI A., WANG S.X. & DUNPHY W.G. 2000. Requirement for Atr in phosphorylation of Chk1 and cell cycle regulation in response to DNA replication blocks and UV-damaged DNA in Xenopus egg extracts. *Genes Dev* 14: 2745-2756.

HABUCHI S., TSUTSUI H., KOCHANIAK A.B., MIYAWAKI A. & VAN OIJEN A.M. 2008. mKikGR, a monomeric photoswitchable fluorescent protein. *PLoS One* 3: e3944.

HARTWELL L.H., CULOTTI J., PRINGLE J.R. & REID B.J. 1974. Genetic control of the cell division cycle in yeast. *Science* 183: 46-51.

HAWKINS M., RETKUTE R., MÜLLER C.A., SANER N., TANAKA T.U., DE MOURA A.P. &

NIEDUSZYNSKI C.A. 2013. High-resolution replication profiles define the stochastic nature of genome replication initiation and termination. *Cell Rep* 5: 1132-1141.

HELLER R.C., KANG S., LAM W.M., CHEN S., CHAN C.S. & BELL S.P. 2011. Eukaryotic origin-dependent DNA replication in vitro reveals sequential action of DDK and S-CDK kinases. *Cell* 146: 80-91.

HENNESSY K.M., CLARK C.D. & BOTSTEIN D. 1990. Subcellular localization of yeast CDC46 varies with the cell cycle. *Genes Dev* 4: 2252-2263.

HEUN P., LAROCHE T., RAGHURAMAN M.K. & GASSER S.M. 2001. The positioning and dynamics of origins of replication in the budding yeast nucleus. *J Cell Biol* 152: 385-400.

HIRAGA S., ALVINO G.M., CHANG F., LIAN H.Y., SRIDHAR A., KUBOTA T., BREWER B.J., WEINREICH M., RAGHURAMAN M.K. & DONALDSON A.D. 2014. Rif1 controls DNA replication by directing Protein Phosphatase 1 to reverse Cdc7-mediated phosphorylation of the MCM complex. *Genes Dev* 28: 372-383.

HIRAGA S., ROBERTSON E.D. & DONALDSON A.D. 2006. The Ctf18 RFC-like complex positions yeast telomeres but does not specify their replication time. *EMBO J* 25: 1505-1514.

HIRATANI I., RYBA T., ITOH M., RATHJEN J., KULIK M., PAPP B., FUSSNER E., BAZETT-JONES D.P., PLATH K., DALTON S., RATHJEN P.D. & GILBERT D.M. 2010. Genome-wide dynamics of replication timing revealed by in vitro models of

mouse embryogenesis. *Genome Res* 20: 155-169.

HO B., BARYSHNIKOVA A. & BROWN G.W. 2018. Unification of Protein Abundance Datasets Yields a Quantitative Saccharomyces cerevisiae Proteome. *Cell Syst* 6: 192-205.e3.

HOGGARD T., SHOR E., MÜLLER C.A., NIEDUSZYNSKI C.A. & FOX C.A. 2013. A Link between ORC-origin binding mechanisms and origin activation time revealed in budding yeast. *PLoS Genet* 9: e1003798.

HONEYCUTT K.A., CHEN Z., KOSTER M.I., MIERS M., NUCHTERN J., HICKS J., ROOP D.R. & SHOHET J.M. 2006. Deregulated minichromosomal maintenance protein MCM7 contributes to oncogene driven tumorigenesis. *Oncogene* 25: 4027-4032.

HU J., SUN L., SHEN F., CHEN Y., HUA Y., LIU Y., ZHANG M., HU Y., WANG Q., XU W., SUN F., JI J., MURRAY J.M., CARR A.M. & KONG D. 2012. The intra-S phase checkpoint targets Dna2 to prevent stalled replication forks from reversing. *Cell* 149: 1221-1232.

HYRIEN O. & MECHALI M. 1993. Chromosomal replication initiates and terminates at random sequences but at regular intervals in the ribosomal DNA of Xenopus early embryos. *EMBO J* 12: 4511-4520.

IBARRA A., SCHWOB E. & MÉNDEZ J. 2008. Excess MCM proteins protect human cells from replicative stress by licensing backup origins of replication. *Proc Natl Acad Sci U S A* 105: 8956-8961.

ILVES I., PETOJEVIC T., PESAVENTO J.J. & BOTCHAN M.R. 2010. Activation of the

MCM2-7 helicase by association with Cdc45 and GINS proteins. *Mol Cell* 37: 247-258.

IVESSA A.S., ZHOU J.Q. & ZAKIAN V.A. 2000. The Saccharomyces Pif1p DNA helicase and the highly related Rrm3p have opposite effects on replication fork progression in ribosomal DNA. *Cell* 100: 479-489.

IYER D.R. & RHIND N. 2013. Checkpoint regulation of replication forks: global or local? *Biochem Soc Trans* 41: 1701-1705.

JASIN M. & ROTHSTEIN R. 2013. Repair of strand breaks by homologous recombination. *Cold Spring Harb Perspect Biol* 5: a012740.

KAGUNI J.M. 2011. Replication initiation at the Escherichia coli chromosomal origin. *Curr Opin Chem Biol* 15: 606-613.

KAMIMURA Y., TAK Y.S., SUGINO A. & ARAKI H. 2001. Sld3, which interacts with Cdc45 (Sld4), functions for chromosomal DNA replication in Saccharomyces cerevisiae. *EMBO J* 20: 2097-2107.

KANG S., WARNER M.D. & BELL S.P. 2014. Multiple functions for Mcm2-7 ATPase motifs during replication initiation. *Mol Cell* 55: 655-665.

KATO H., MIYAZAKI T., FUKAI Y., NAKAJIMA M., SOHDA M., TAKITA J., MASUDA N., FUKUCHI M., MANDA R., OJIMA H., TSUKADA K., ASAO T. & KUWANO H. 2003. A new proliferation marker, minichromosome maintenance protein 2, is associated with tumor aggressiveness in esophageal squamous cell carcinoma. *J Surg Oncol* 84: 24-30.

KEPPLER A., GENDREIZIG S., GRONEMEYER T., PICK H., VOGEL H. & JOHNSSON K.

2003. A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat Biotechnol* 21: 86-89.

KIM H.S., NICKOLOFF J.A., WU Y., WILLIAMSON E.A., SIDHU G.S., REINERT B.L., JAISWAL A.S., SRINIVASAN G., PATEL B., KONG K., BURMA S., LEE S.H. & HROMAS R.A. 2017. Endonuclease EEPD1 Is a Gatekeeper for Repair of Stressed Replication Forks. *J Biol Chem* 292: 2795-2804.

KNOTT S.R., PEACE J.M., OSTROW A.Z., GAN Y., REX A.E., VIGGIANI C.J., TAVARÉ S. & APARICIO O.M. 2012. Forkhead transcription factors establish origin timing and long-range clustering in S. cerevisiae. *Cell* 148: 99-111.

KOBAYASHI T. & HORIUCHI T. 1996. A yeast gene product, Fob1 protein, required for both replication fork blocking and recombinational hotspot activities. *Genes Cells* 1: 465-474.

KRYSAN P.J. & CALOS M.P. 1991. Replication initiates at multiple locations on an autonomously replicating plasmid in human cells. *Mol Cell Biol* 11: 1464-1472.

KUMAR S. & HUBERMAN J.A. 2009. Checkpoint-dependent regulation of origin firing and replication fork movement in response to DNA damage in fission yeast. *Mol Cell Biol* 29: 602-611.

KUNNEV D., FREELAND A., QIN M., LEACH R.W., WANG J., SHENOY R.M. & PRUITT S.C. 2015. Effect of minichromosome maintenance protein 2 deficiency on the locations of DNA replication origins. *Genome Res* 25: 558-569.

KUNNEV D., RUSINIAK M.E., KUDLA A., FREELAND A., CADY G.K. & PRUITT S.C. 2010.

DNA damage response and tumorigenesis in Mcm2-deficient mice. *Oncogene* 29: 3630-3638.

LABIB K. & DE PICCOLI G. 2011. Surviving chromosome replication: the many roles of the S-phase checkpoint pathway. *Philos Trans R Soc Lond B Biol Sci* 366: 3554-3561.

LABIB K., DIFFLEY J.F. & KEARSEY S.E. 1999. G1-phase and B-type cyclins exclude the DNA-replication factor Mcm4 from the nucleus. *Nat Cell Biol* 1: 415-422.

LABIB K., KEARSEY S.E. & DIFFLEY J.F. 2001. MCM2-7 proteins are essential components of prereplicative complexes that accumulate cooperatively in the nucleus during G1-phase and are required to establish, but not maintain, the S-phase checkpoint. *Mol Biol Cell* 12: 3658-3667.

LANG G.I. & MURRAY A.W. 2011. Mutation rates across budding yeast chromosome VI are correlated with replication timing. *Genome Biol Evol* 3: 799-811.

LANGSTON L.D. & O'DONNELL M.E. 2019. An explanation for origin unwinding in eukaryotes. *Elife* 8:

LANGSTON L.D., ZHANG D., YURIEVA O., GEORGESCU R.E., FINKELSTEIN J., YAO N.Y., INDIANI C. & O'DONNELL M.E. 2014. CMG helicase and DNA polymerase epsilon form a functional 15-subunit holoenzyme for eukaryotic leading-strand DNA replication. *Proc Natl Acad Sci U S A* 111: 15390-15395.

LARSON K., SAHM J., SHENKAR R. & STRAUSS B. 1985. Methylation-induced blocks to in vitro DNA replication. *Mutat Res* 150: 77-84.

LEBOFSKY R., TAKAHASHI T. & WALTER J.C. 2009. DNA replication in nucleus-free Xenopus egg extracts. *Methods Mol Biol* 521: 229-252.

LEI M., KAWASAKI Y. & TYE B.K. 1996. Physical interactions among Mcm proteins and effects of Mcm dosage on DNA replication in Saccharomyces cerevisiae. *Mol Cell Biol* 16: 5081-5090.

LIAN H.Y., ROBERTSON E.D., HIRAGA S., ALVINO G.M., COLLINGWOOD D., MCCUNE H.J., SRIDHAR A., BREWER B.J., RAGHURAMAN M.K. & DONALDSON A.D. 2011. The effect of Ku on telomere replication time is mediated by telomere length but is independent of histone tail acetylation. *Mol Biol Cell* 22: 1753-1765.

LIANG D.T., HODSON J.A. & FORSBURG S.L. 1999. Reduced dosage of a single fission yeast MCM protein causes genetic instability and S phase delay. *J Cell Sci* 112: 559-567.

LIKU M.E., NGUYEN V.Q., ROSALES A.W., IRIE K. & LI J.J. 2005. CDK phosphorylation of a novel NLS-NES module distributed between two subunits of the Mcm2-7 complex prevents chromosomal rereplication. *Mol Biol Cell* 16: 5026-5039.

LIPFORD J.R. & BELL S.P. 2001. Nucleosomes positioned by ORC facilitate the initiation of DNA replication. *Mol Cell* 7: 21-30.

LIU B., BASKIN R.J. & KOWALCZYKOWSKI S.C. 2013. DNA unwinding heterogeneity by RecBCD results from static molecules able to equilibrate. *Nature* 500: 482-485.

LOEWEN P.C. & SWITALA J. 1995. Template secondary structure can increase the

error frequency of the DNA polymerase from Thermus aquaticus. *Gene* 164: 59-63.

LOPES M., COTTA-RAMUSINO C., PELLICIOLI A., LIBERI G., PLEVANI P., MUZI-FALCONI M., NEWLON C.S. & FOIANI M. 2001. The DNA replication checkpoint response stabilizes stalled replication forks. *Nature* 412: 557-561.

LOSSAINT G., LARROQUE M., RIBEYRE C., BEC N., LARROQUE C., DECAILLET C., GARI K. & CONSTANTINOU A. 2013. FANCD2 binds MCM proteins and controls replisome function upon activation of s phase checkpoint signaling. *Mol Cell* 51: 678-690.

LOVELAND A.B., HABUCHI S., WALTER J.C. & VAN OIJEN A.M. 2012. A general approach to break the concentration barrier in single-molecule imaging. *Nat Methods* 9: 987-992.

MACALPINE H.K., GORDÂN R., POWELL S.K., HARTEMINK A.J. & MACALPINE D.M. 2010. Drosophila ORC localizes to open chromatin and marks sites of cohesin complex loading. *Genome Res* 20: 201-211.

MAHBUBANI H.M., CHONG J.P., CHEVALIER S., THÖMMES P. & BLOW J.J. 1997. Cell cycle regulation of the replication licensing system: involvement of a Cdk-dependent inhibitor. *J Cell Biol* 136: 125-135.

MAHBUBANI H.M., PAULL T., ELDER J.K. & BLOW J.J. 1992. DNA replication initiates at multiple sites on plasmid DNA in Xenopus egg extracts. *Nucleic Acids Res* 20: 1457-1462.

MANTIERO D., MACKENZIE A., DONALDSON A. & ZEGERMAN P. 2011. Limiting

replication initiation factors execute the temporal programme of origin firing in budding yeast. *EMBO J* 30: 4805-4814.

MARAHRENS Y. & STILLMAN B. 1992. A yeast chromosomal origin of DNA replication defined by multiple functional elements. *Science* 255: 817-823.

MARIC M., MACULINS T., DE PICCOLI G. & LABIB K. 2014. Cdc48 and a ubiquitin ligase drive disassembly of the CMG helicase at the end of DNA replication. *Science* 346: 1253596.

MATTAROCCI S., SHYIAN M., LEMMENS L., DAMAY P., ALTINTAS D.M., SHI T., BARTHOLOMEW C.R., THOMÄ N.H., HARDY C.F. & SHORE D. 2014. Rif1 controls DNA replication timing in yeast through the PP1 phosphatase Glc7. *Cell Rep* 7: 62-69.

MCGUFFEE S.R., SMITH D.J. & WHITEHOUSE I. 2013. Quantitative, genome-wide analysis of eukaryotic replication initiation and termination. *Mol Cell* 50: 123-135.

MERRICK C.J., JACKSON D. & DIFFLEY J.F. 2004. Visualization of altered replication dynamics after DNA damage in human cells. *J Biol Chem* 279: 20067-20075.

MILLER T.C.R., LOCKE J., GREIWE J.F., DIFFLEY J.F.X. & COSTA A. 2019. Mechanism of head-to-head MCM double-hexamer formation revealed by cryo-EM. *Nature* 575: 704-710.

MOISEEVA T.N. & BAKKENIST C.J. 2018. Regulation of the initiation of DNA replication in human cells. *DNA Repair (Amst)* 72: 99-106.

MORTIMER R.K. 2000. Evolution and variation of the yeast (Saccharomyces) genome. *Genome Res* 10: 403-409.

MOYER S.E., LEWIS P.W. & BOTCHAN M.R. 2006. Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc Natl Acad Sci U S A* 103: 10236-10241.

MÜLLER C.A., HAWKINS M., RETKUTE R., MALLA S., WILSON R., BLYTHE M.J., NAKATO R., KOMATA M., SHIRAHIGE K., DE MOURA A.P. & NIEDUSZYNSKI C.A. 2014. The dynamics of genome replication using deep sequencing. *Nucleic Acids Res* 42: e3.

MÜLLER C.A. & NIEDUSZYNSKI C.A. 2017. DNA replication timing influences gene expression level. *J Cell Biol* 216: 1907-1914.

MURAMATSU S., HIRAI K., TAK Y.S., KAMIMURA Y. & ARAKI H. 2010. CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol (epsilon}, and GINS in budding yeast. *Genes Dev* 24: 602-612.

NATSUME T., MÜLLER C.A., KATOU Y., RETKUTE R., GIERLIŃSKI M., ARAKI H., BLOW J.J., SHIRAHIGE K., NIEDUSZYNSKI C.A. & TANAKA T.U. 2013. Kinetochores coordinate pericentromeric cohesion and early DNA replication by Cdc7-Dbf4 kinase recruitment. *Mol Cell* 50: 661-674.

NGUYEN V.Q., CO C., IRIE K. & LI J.J. 2000. Clb/Cdc28 kinases promote nuclear export of the replication initiator proteins Mcm2-7. *Curr Biol* 10: 195-205.

NGUYEN V.Q., CO C. & LI J.J. 2001. Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. *Nature* 411: 1068-1073.

NICK MCELHINNY S.A., GORDENIN D.A., STITH C.M., BURGERS P.M. & KUNKEL T.A. 2008. Division of labor at the eukaryotic replication fork. *Mol Cell* 30: 137-144.

NIEDUSZYNSKI C.A., HIRAGA S., AK P., BENHAM C.J. & DONALDSON A.D. 2007. OriDB: a DNA replication origin database. *Nucleic Acids Res* 35: D40-6.

NISHIMURA K., FUKAGAWA T., TAKISAWA H., KAKIMOTO T. & KANEMAKI M. 2009. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat Methods* 6: 917-922.

NISHIMURA K. & KANEMAKI M.T. 2014. Rapid Depletion of Budding Yeast Proteins via the Fusion of an Auxin-Inducible Degron (AID). *Curr Protoc Cell Biol* 64: 20.9.1-16.

O'BRIEN K.P., REMM M. & SONNHAMMER E.L. 2005. Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Res* 33: D476-80.

OHOUO P.Y., BASTOS DE OLIVEIRA F.M., ALMEIDA B.S. & SMOLKA M.B. 2010. DNA damage signaling recruits the Rtt107-Slx4 scaffolds via Dpb11 to mediate replication stress response. *Mol Cell* 39: 300-306.

PADOVAN-MERHAR O., NAIR G.P., BIAESCH A.G., MAYER A., SCARFONE S., FOLEY S.W., WU A.R., CHURCHMAN L.S., SINGH A. & RAJ A. 2015. Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. *Mol Cell* 58: 339-352.

PAINTER R.B. & YOUNG B.R. 1980. Radiosensitivity in ataxia-telangiectasia: a new

explanation. *Proc Natl Acad Sci U S A* 77: 7315-7317.

PAULOVICH A.G. & HARTWELL L.H. 1995. A checkpoint regulates the rate of progression through S phase in S. cerevisiae in response to DNA damage. *Cell* 82: 841-847.

PELLEGRINI L. 2012. The Pol α-primase complex. *Subcell Biochem* 62: 157-169.

POWELL S.K., MACALPINE H.K., PRINZ J.A., LI Y., BELSKY J.A. & MACALPINE D.M. 2015. Dynamic loading and redistribution of the Mcm2-7 helicase complex through the cell cycle. *EMBO J* 34: 531-543.

PURSELL Z.F., ISOZ I., LUNDSTRÖM E.B., JOHANSSON E. & KUNKEL T.A. 2007. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science* 317: 127-130.

RAGHURAMAN M.K., WINZELER E.A., COLLINGWOOD D., HUNT S., WODICKA L., CONWAY A., LOCKHART D.J., DAVIS R.W., BREWER B.J. & FANGMAN W.L. 2001. Replication dynamics of the yeast genome. *Science* 294: 115-121.

REAPER P.M., GRIFFITHS M.R., LONG J.M., CHARRIER J.D., MACCORMICK S., CHARLTON P.A., GOLEC J.M. & POLLARD J.R. 2011. Selective killing of ATM- or p53-deficient cancer cells through inhibition of ATR. *Nat Chem Biol* 7: 428-430.

REMUS D., BEURON F., TOLUN G., GRIFFITH J.D., MORRIS E.P. & DIFFLEY J.F. 2009. Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell* 139: 719-730.

RHIND N. & GILBERT D.M. 2013. DNA replication timing. *Cold Spring Harb*

Perspect Biol 5: a010132.

RYBA T., BATTAGLIA D., CHANG B.H., SHIRLEY J.W., BUCKLEY Q., POPE B.D., DEVIDAS M., DRUKER B.J. & GILBERT D.M. 2012. Abnormal developmental control of replication-timing domains in pediatric acute lymphoblastic leukemia. *Genome Res* 22: 1833-1844.

SALOMON W.E., JOLLY S.M., MOORE M.J., ZAMORE P.D. & SEREBROV V. 2016. Single-Molecule Imaging Reveals that Argonaute Reshapes the Binding Properties of Its Nucleic Acid Guides. *Cell* 166: 517-520.

SANCAR A., LINDSEY-BOLTZ L.A., UNSAL-KACMAZ K. & LINN S. 2004. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu Rev Biochem* 73: 39-85.

SASAKI T., RIVERA-MULIA J.C., VERA D., ZIMMERMAN J., DAS S., PADGET M., NAKAMICHI N., CHANG B.H., TYNER J., DRUKER B.J., WENG A.P., CIVIN C.I., EAVES C.J. & GILBERT D.M. 2017. Stability of patient-specific features of altered DNA replication timing in xenografts of primary human acute lymphoblastic leukemia. *Exp Hematol* 51: 71-82.e3.

SCHRADER C., JANSSEN D., KLAPPER W., SIEBMANN J.U., MEUSERS P., BRITTINGER G., KNEBA M., TIEMANN M. & PARWARESCH R. 2005. Minichromosome maintenance protein 6, a proliferation marker superior to Ki-67 and independent predictor of survival in patients with mantle cell lymphoma. *Br J Cancer* 93: 939-945.

SEGURADO M., DE LUIS A. & ANTEQUERA F. 2003. Genome-wide distribution of

DNA replication origins at A+T-rich islands in Schizosaccharomyces pombe. *EMBO Rep* 4: 1048-1053.

SEILER J.A., CONTI C., SYED A., ALADJEM M.I. & POMMIER Y. 2007. The intra-S-phase checkpoint affects both DNA replication initiation and elongation: single-cell and -DNA fiber analyses. *Mol Cell Biol* 27: 5806-5818.

SHEN Z., CHAKRABORTY A., JAIN A., GIRI S., HA T., PRASANTH K.V. & PRASANTH S.G. 2012. Dynamic association of ORCA with prereplicative complex components regulates DNA replication initiation. *Mol Cell Biol* 32: 3107-3120.

SHEU Y.J. & STILLMAN B. 2006. Cdc7-Dbf4 phosphorylates MCM proteins via a docking site-mediated mechanism to promote S phase progression. *Mol Cell* 24: 101-113.

SHU D., ZHANG H., JIN J. & GUO P. 2007. Counting of six pRNAs of phi29 DNA-packaging motor with customized single-molecule dual-view system. *EMBO J* 26: 527-537.

SIMPSON R.T. 1990. Nucleosome positioning can affect the function of a cis-acting DNA element in vivo. *Nature* 343: 387-389.

SMITH C.S., JOURAVLEVA K., HUISMAN M., JOLLY S.M., ZAMORE P.D. & GRUNWALD D. 2019. An automated Bayesian pipeline for rapid analysis of single-molecule binding data. *Nat Commun* 10: 272.

SMITH D.J. & WHITEHOUSE I. 2012. Intrinsic coupling of lagging-strand synthesis to chromatin assembly. *Nature* 483: 434-438.

SOGO J.M., LOPES M. & FOIANI M. 2002. Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects. *Science* 297: 599-602.

SORIANO I., MORAFRAILE E.C., VÁZQUEZ E., ANTEQUERA F. & SEGURADO M. 2014. Different nucleosomal architectures at early and late replicating origins in Saccharomyces cerevisiae. *BMC Genomics* 15: 791.

SPECK C., CHEN Z., LI H. & STILLMAN B. 2005. ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA. *Nat Struct Mol Biol* 12: 965-971.

STANOJCIC S., LEMAITRE J.M., BRODOLIN K., DANIS E. & MECHALI M. 2008. In Xenopus egg extracts, DNA replication initiates preferentially at or near asymmetric AT sequences. *Mol Cell Biol* 28: 5265-5274.

STINCHCOMB D.T., STRUHL K. & DAVIS R.W. 1979. Isolation and characterisation of a yeast chromosomal replicator. *Nature* 282: 39-43.

TAKARA T.J. & BELL S.P. 2011. Multiple Cdt1 molecules act at each origin to load replication-competent Mcm2-7 helicases. *EMBO J* 30: 4885-4896.

TANAKA S. & DIFFLEY J.F. 2002. Interdependent nuclear accumulation of budding yeast Cdt1 and Mcm2-7 during G1 phase. *Nat Cell Biol* 4: 198-207.

TANAKA S., NAKATO R., KATOU Y., SHIRAHIGE K. & ARAKI H. 2011. Origin association of Sld3, Sld7, and Cdc45 proteins is a key step for determination of origin-firing timing. *Curr Biol* 21: 2055-2063.

TANAKA S., UMEMORI T., HIRAI K., MURAMATSU S., KAMIMURA Y. & ARAKI H. 2007. CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. *Nature* 445: 328-332.

TERCERO J.A. & DIFFLEY J.F. 2001. Regulation of DNA replication fork progression through damaged DNA by the Mec1/Rad53 checkpoint. *Nature* 412: 553-557.

TERCERO J.A., LONGHESE M.P. & DIFFLEY J.F. 2003. A central role for DNA replication forks in checkpoint activation and response. *Mol Cell* 11: 1323-1336.

THOMAS D.C., ROBERTS J.D., SABATINO R.D., MYERS T.W., TAN C.K., DOWNEY K.M., SO A.G., BAMBARA R.A. & KUNKEL T.A. 1991. Fidelity of mammalian DNA replication and replicative DNA polymerases. *Biochemistry* 30: 11751-11759.

TICAU S., FRIEDMAN L.J., CHAMPASA K., CORRÊA I.R., GELLES J. & BELL S.P. 2017. Mechanism and timing of Mcm2-7 ring closure during DNA replication origin licensing. *Nat Struct Mol Biol* 24: 309-315.

TICAU S., FRIEDMAN L.J., IVICA N.A., GELLES J. & BELL S.P. 2015. Single-Molecule Studies of Origin Licensing Reveal Mechanisms Ensuring Bidirectional Helicase Loading. *Cell*

ULBRICH M.H. & ISACOFF E.Y. 2007. Subunit counting in membrane-bound proteins. *Nat Methods* 4: 319-321.

UNNIKRISHNAN A., GAFKEN P.R. & TSUKIYAMA T. 2010. Dynamic changes in histone acetylation regulate origins of DNA replication. *Nat Struct Mol Biol* 17: 430-437.

VAN DEURSEN F., SENGUPTA S., DE PICCOLI G., SANCHEZ-DIAZ A. & LABIB K. 2012.

Mcm10 associates with the loaded DNA helicase at replication origins and defines a novel step in its activation. *EMBO J* 31: 2195-2206.

VISNAPUU M.L., DUZDEVICH D. & GREENE E.C. 2008. The importance of surfaces in single-molecule bioscience. *Mol Biosyst* 4: 394-403.

VOGELAUER M., RUBBI L., LUCAS I., BREWER B.J. & GRUNSTEIN M. 2002. Histone acetylation regulates the time of replication origin firing. *Mol Cell* 10: 1223-1233.

VOICHEK Y., BAR-ZIV R. & BARKAI N. 2016. Expression homeostasis during DNA replication. *Science* 351: 1087-1090.

WAL M. & PUGH B.F. 2012. Genome-wide mapping of nucleosome positions in yeast using high-resolution MNase ChIP-Seq. *Methods Enzymol* 513: 233-250.

WALTER J., SUN L. & NEWPORT J. 1998. Regulated chromosomal DNA replication in the absence of a nucleus. *Mol Cell* 1: 519-529.

WARNER M.D., AZMI I.F., KANG S., ZHAO Y. & BELL S.P. 2017. Replication origin-flanking roadblocks reveal origin-licensing dynamics and altered sequence dependence. *J Biol Chem* 292: 21417-21430.

WASSERMAN M.R., SCHAUER G.D., O'DONNELL M.E. & LIU S. 2019. Replication Fork Activation Is Enabled by a Single-Stranded DNA Gate in CMG Helicase. *Cell* 178: 600-611.e16.

WATASE G., TAKISAWA H. & KANEMAKI M.T. 2012. Mcm10 plays a role in functioning of the eukaryotic replicative DNA helicase, Cdc45-Mcm-GINS.

Curr Biol 22: 343-349.

WILLIAMSON D.H. 1965. The timing of deoxyribonucleic acid synthesis in the cell cycle of Saccharomyces cerevisiae. *J Cell Biol* 25: 517-528.

WILMES G.M. & BELL S.P. 2002. The B2 element of the Saccharomyces cerevisiae ARS1 origin of replication requires specific sequences to facilitate pre-RC formation. *Proc Natl Acad Sci U S A* 99: 101-106.

WOODWARD A.M., GÖHLER T., LUCIANI M.G., OEHLMANN M., GE X., GARTNER A., JACKSON D.A. & BLOW J.J. 2006. Excess Mcm2-7 license dormant origins of replication that can be used under conditions of replicative stress. *J Cell Biol* 173: 673-683.

WU R., WANG J. & LIANG C. 2012. Cdt1p, through its interaction with Mcm6p, is required for the formation, nuclear accumulation and chromatin loading of the MCM complex. *J Cell Sci* 125: 209-219.

WYATT M.D. & PITTMAN D.L. 2006. Methylating agents and DNA repair responses: Methylated bases and sources of strand breaks. *Chem Res Toxicol* 19: 1580-1594.

XU W., APARICIO J.G., APARICIO O.M. & TAVARÉ S. 2006. Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in S. cerevisiae. *BMC Genomics* 7: 276.

YAN H., MERCHANT A.M. & TYE B.K. 1993. Cell cycle-regulated nuclear localization of MCM2 and MCM3, which are required for the initiation of DNA synthesis

at chromosomal replication origins in yeast. *Genes Dev* 7: 2149-2160.

YANG S.C., RHIND N. & BECHHOEFER J. 2010. Modeling genome-wide replication kinetics reveals a mechanism for regulation of replication timing. *Mol Syst Biol* 6: 404.

YARDIMCI H., LOVELAND A.B., HABUCHI S., VAN OIJEN A.M. & WALTER J.C. 2010. Uncoupling of sister replisomes during eukaryotic DNA replication. *Mol Cell* 40: 834-840.

YARDIMCI H., LOVELAND A.B., VAN OIJEN A.M. & WALTER J.C. 2012. Single-molecule analysis of DNA replication in Xenopus egg extracts. *Methods* 57: 179-186.

YARDIMCI H. & WALTER J.C. 2014. Prereplication-complex formation: a molecular double take. *Nat Struct Mol Biol* 21: 20-25.

YEELES J.T., DEEGAN T.D., JANSKA A., EARLY A. & DIFFLEY J.F. 2015. Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature* 519: 431-435.

YOKOTA H., CHUJO Y.A. & HARADA Y. 2013. Single-molecule imaging of the oligomer formation of the nonhexameric Escherichia coli UvrD helicase. *Biophys J* 104: 924-933.

YOSHIDA K., BACAL J., DESMARAIS D., PADIOLEAU I., TSAPONINA O., CHABES A., PANTESCO V., DUBOIS E., PARRINELLO H., SKRZYPCZAK M., GINALSKI K., LENGRONNE A. & PASERO P. 2014. The histone deacetylases sir2 and rpd3 act on ribosomal DNA to control the replication program in budding yeast.

Mol Cell 54: 691-697.

YU Y., ULBRICH M.H., LI M.H., DOBBINS S., ZHANG W.K., TONG L., ISACOFF E.Y. & YANG J. 2012. Molecular mechanism of the assembly of an acid-sensing receptor ion channel complex. *Nat Commun* 3: 1252.

ZEGERMAN P. & DIFFLEY J.F. 2007. Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature* 445: 281-285.

ZENTNER G.E., KASINATHAN S., XIN B., ROHS R. & HENIKOFF S. 2015. ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo. *Nat Commun* 6: 8733.

ZHOU Z.X., LUJAN S.A., BURKHOLDER A.B., GARBACZ M.A. & KUNKEL T.A. 2019. Roles for DNA polymerase δ in initiating and terminating leading strand DNA replication. *Nat Commun* 10: 3992.

ZOU L. & STILLMAN B. 2000. Assembly of a complex containing Cdc45p, replication protein A, and Mcm2p at replication origins controlled by S-phase cyclin-dependent kinases and Cdc7p-Dbf4p kinase. *Mol Cell Biol* 20: 3086-3096.